Applied Economics, 2008, **40**, 3101–3110

Asymptotic bias reduction for a conditional marginal effects estimator in sample selection models

Alpaslan Akay* and Elias Tsakas

Department of Economics, University of Göteborg, P.O. Box 640, 405 30 Göteborg, Sweden

In this article we discuss the differences between the average marginal effect and the marginal effect of the average individual in sample selection models, estimated by the Heckman procedure. We show that the bias that emerges as a consequence of interchanging the measures, could be very significant, even in the limit. We suggest a computationally cheap approximation method, which corrects the bias to a large extent. We illustrate the implications of our method with an empirical application of earnings assimilation and a small Monte Carlo simulation.

I. Introduction

A large amount of applied work using nonlinear microeconometric models has been carried out over the last few decades. One of the important characteristics of these models is their nature, which allows the calculation of individual marginal effects. In general, most empirical studies report one of the two established point estimators for marginal effects: (i) the average of the marginal effects of all individuals in the sample, and (ii) the marginal effect at the sample means. Neglecting their quantitative, and more importantly, conceptual differences is a quite common practice. Greene's (2003) discussion on the marginal effects in binary choice models stresses the fact that in many occasions the asymptotic equivalence of the two measures is taken for granted. Verlinda (2006) shows that arbitrarily interchanging them in a binary pro-bit model could create bias and lead to misleading conclusions, since the two measures estimate different quantities

In the present article we discuss the relationship between the two measures in the context of sample selection models, also known as Tobit type II (Heckman, 1976, 1979). Provided that one is interested in the average effect over the population rather than in the effect over the average individual, we show that evaluating the derivative at the sample means leads to biased predictions, even asymptotically. Since the other alternative (averaging the marginal effects for the whole sample) could be computationally inefficient, we propose an approximation technique which significantly reduces the bias, without significantly increasing the number of numerical operations. In order to accomplish this, we express the average marginal effect (*AME*) with the Taylor expansion around the mean values of the explanatory variables and prove that the conventionally used marginal effect of the average individual (*MEAI*) is actually equal to the first order Taylor approximation, while the order of magnitude is equal to the asymptotic bias. By shifting to the second order

*Corresponding author. E-mail: Alpaslan.Akay@economics.gu.se

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2008 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/00036840600994096

approximation, one can reduce the size of the bias without high computational cost, since the second term of the series is a function of the Hessian and the covariance matrix evaluated at the sample means.

Marginal effects in sample selection models have recently been discussed. Saha *et al.* (1997a) show that failure to account for changes in the inverse of Mill's ratio leads to biased marginal effects. Hoffmann and Kassouf (2005) introduce unconditional marginal effects in addition to the standard conditional ones. In any case, the clear distinction between *AME* and *MEAI* is necessary regardless of the definition of the marginal effects.

In order to emphasize the necessity of a consistent estimator for the *AME* we present an empirical application of immigrant earnings assimilation using registered data from Sweden. We find that our approach corrects the bias to a large extent, and discuss the policy implications behind this relative difference.

The article has the following structure. Section II briefly describes Heckman's two step procedure. Section III introduces the theoretical results of our approach. In Section IV we apply the model to real data, and also include Monte Carlo simulations. Section V concludes the article.

II. Heckman Procedure and Marginal Effects

Consider the following sample selection (otherwise known as the Tobit type II) model:

$$Y_i^* = \mathbf{X}_i'\beta + \varepsilon_i \tag{1}$$

$$H_i^* = \mathbf{Z}_i\gamma + u_i \tag{2}$$

$$H_i = 1\left[H_i^* > 0\right] \tag{3}$$

$$Y_i = Y_i^* \cdot H_i \tag{4}$$

where $i = 1, \ldots, N$. Let the latent variables Y_i^* and H_i^* denote individual i's earnings and hours of work respectively. Assume also that the matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ include various observed individual characteristics, with $\mathbf{X}_i$ being a strict subset of $\mathbf{Z}_i$. Finally, the joint error term $(\varepsilon_i,\ u_i)$ follows the bivariate normal distribution with correlation coefficient ρ and normalized variance of the selection equation error term, $\sigma_u^2 = 1$. Our primary aim is to estimate the parameter vector β of the earnings equation. We know that strictly positive hours of work is a necessary and sufficient condition for participating in the job market, ie. $H_i^* > 0$. Then the participation decision takes the form of a binary choice, since *working* and *not working* are complementary events, and as such they can be written as the indicator function of the equation above.

Conditioning on the subset of the population that contains the individuals who actually work, the expectation of the earnings given participation would be given by the following formula (Greene, 2003):

$$\begin{aligned} E\left[Y_i^*|H_i = 1, \mathbf{X}_i, \mathbf{Z}_i\right] &= E\left[\mathbf{X}_i'\beta + \varepsilon_i|H_i^* > 0\right] \\ &= \mathbf{X}_i'\hat{\beta} + E\left[\varepsilon_i|u_i > -\mathbf{Z}_i'\gamma\right] \\ &= \mathbf{X}_i'\hat{\beta} + \hat{\rho}\hat{\sigma}_\varepsilon \frac{\phi\left(-\mathbf{Z}_i'\hat{\gamma}\right)}{1 - \Phi\left(-\mathbf{Z}_i'\hat{\gamma}\right)} \end{aligned} \tag{5}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and the cumulative distribution of the standard normal distribution respectively. After some notation simplification Equation 5 is rewritten as follows:

$$E\left[Y_i^*|H_i = 1, \mathbf{X}_i, \mathbf{Z}_i\right] = \mathbf{X}_i'\hat{\beta} + \hat{\rho}\hat{\sigma}_\varepsilon\hat{\lambda}_i(\hat{\alpha}_u) \tag{6}$$

where $\hat{\alpha}_u = -\mathbf{Z}_i'\hat{\gamma}$, while λ denotes the inverse of Mill's ratio, ie. $\lambda = \phi/(1 - \Phi)$. It is straightforward that Equation 6 cannot be estimated consistently with ordinary least squares (OLS) in the existence of correlation between ε_i and u_i $(\rho \neq 0)$. On the other hand, although consistent, the maximum likelihood estimator (MLE) constitutes a computationally challenging task. Heckman (1976) introduced a method which can simultaneously handle consistency and computational efficiency. His procedure consists of two separate steps. First, estimate the participation probability by applying a binary probit model

$$P[H_i = 1|\mathbf{Z}_i] = \Phi(\mathbf{Z}_i'\gamma) \tag{7}$$

and use the estimated choice probabilities to calculate $\hat{\lambda}_i(\hat{\alpha}_u)$. In the second step, apply OLS on the earnings equation, while perceiving the estimated inverse Mill's ratio as another explanatory variable. Thus, one gets rid of the omitted variable problem that would otherwise emerge, and the estimator of the parameter vector in the target equation becomes consistent.

The *ceteris paribus* estimated marginal effect[1] of an infinitesimal change of an arbitrary individual characteristic k on individual i's earnings is given

[1] A more precise terminology would require defining it as *conditional marginal effect*, since it refers only to the individuals who actually work.

by the following equation for an explanatory variable $x_{k,i}$:

$$\widehat{ME}_{k,i} = \frac{\partial E[Y_i^* | H_i = 1, \mathbf{X}_i, \mathbf{Z}_i]}{\partial X_{k,i}} = \hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \hat{\delta}_i(\hat{\alpha}_u) \quad (8)$$

where $\hat{\delta}_i(\hat{\alpha}_u) = \hat{\lambda}_i^2(\hat{\alpha}_u) - \hat{\alpha}_u \hat{\lambda}_i(\hat{\alpha}_u)$. The (total) marginal effect of a variable in a sample selection model can be separated into two parts (Greene, 2003). The *direct effect* ($\hat{\beta}_k$) shows the marginal effect of an explanatory variable on the earnings without taking into account the effect of selectivity in the data. The second term in Equation 8 is called *indirect effect* and is a function of the observed individual characteristics. Due to this functional relationship, marginal effects vary across individuals. Omitting the indirect effect would linearize the marginal effect, which is rather convenient in practical terms, but it also creates nonnegligible bias. Such a problem would not arise if the estimated correlation coefficient between the errors of the first and second stage estimation equations (ρ) were equal to zero (Saha *et al*., 1997a).

Since policy decisions upon an action that changes an explanatory variable affecting the whole population, the existence of such nonlinearity allows the use of different measures for the marginal effects. In general, economists are interested in the AME of this action over all individuals. Using an inconsistent estimator for the AME could therefore potentially lead to wrong conclusions and undesired effects of the policy application. A consistent estimator for AME is given by the following expression:

$$\widehat{AME}_k = \frac{1}{N} \sum_{i=1}^{N} \widehat{ME}_{k,i} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \hat{\delta}_i(\hat{\alpha}_u) \right) \quad (9)$$

This follows directly from Khinchine's weak law of large numbers. Namely,

$$\text{plim}_{N \to \infty} \widehat{AME}_k = E\left[\hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \hat{\delta}_i(\alpha_u) \right] = \hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon E\left[\hat{\delta}_i(\hat{\alpha}_u) \right] \quad (10)$$

for every k.

However, due to factors such as computational inefficiency or unavailability of software routines for the calculation of $\widehat{AME}$, researchers usually report the marginal effect of the average individual ($\widehat{MEAI}$), which is equivalent to evaluating the marginal effects at the sample means:

$$\widehat{MEAI}_k = \widehat{ME}_{k,i}|_{\mathbf{Z}_i = \bar{\mathbf{Z}}, \mathbf{X}_i = \bar{\mathbf{X}}} = \hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \bar{\delta} \quad (11)$$

where $\bar{\delta} = \hat{\delta}_i(-\bar{\mathbf{Z}}'\hat{\gamma})$. Notice that $\widehat{MEAI}$ is a consistent estimator for its population counterpart ($MEAI$),

$$\text{plim}_{N \to \infty} \widehat{MEAI}_k = E\left[\hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \frac{\hat{\sigma}_\varepsilon}{\hat{\sigma}_u} \hat{\delta}_i(\bar{\mathbf{Z}}'\hat{\gamma}) \right] = \hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \hat{\delta}_i(\mathbf{M}'\hat{\gamma}) \quad (12)$$

but not for the AME, since $E[\hat{\delta}_i(\hat{\alpha}_u)] \neq \hat{\delta}_i(\mathbf{M}'\hat{\gamma})$. That is, $\widehat{AME}$ and $\widehat{MEAI}$ not only differ quantitatively, but also conceptually, since they estimate different things. Hence, the researcher who arbitrarily interchanges them could be led to misleading conclusions.

III. Approximating Average Marginal Effects

As we discussed above, interchanging $\widehat{AME}$ and $\widehat{MEAI}$ produces bias and leads to inconsistent estimation of AME. In this section we suggest an approximation method for estimating AME that is computationally efficient and that significantly reduces the bias emerging from the use of $\widehat{MEAI}$. In order to extract the asymptotic bias we expand the Taylor series of $\hat{\delta}_i(\mathbf{Z}_i'\hat{\gamma})$ around the mean of the explanatory variables, $\mathbf{M}$:

$$\begin{aligned}\hat{\delta}_i(\mathbf{Z}'_i\hat{\gamma}) &= \hat{\delta}_i(\mathbf{M}'\hat{\gamma}) + \sum_k \left(\left.\frac{\partial \hat{\delta}_i(\mathbf{Z}_i'\hat{\gamma})}{\partial Z_k}\right|_{\mathbf{M}} \cdot (Z_{k,i} - M_k) \right) \\ &+ \frac{1}{2!} \sum_{k_1} \sum_{k_2} \left(\left.\frac{\partial^2 \hat{\delta}_i(\mathbf{Z}_i'\hat{\gamma})}{\partial Z_{k_1,i} \partial Z_{k_2,i}}\right|_{\mathbf{M}} \cdot (Z_{k_1,i} - M_{k_1})(Z_{k_2,i} - M_{k_2}) \right) \\ &+ \ldots = \hat{\delta}_i(\mathbf{M}'\hat{\gamma}) \\ &+ \sum_{j=1}^{\infty} \left[\frac{1}{j!} \sum_{k_1,\ldots,k_j} \left(\left.\frac{\partial^j \hat{\delta}_i(\mathbf{Z}_i'\hat{\gamma})}{\partial Z_{k_1 i}, \ldots, \partial Z_{k_j,i}}\right|_{\mathbf{M}} \cdot (Z_{k_1,i} - M_{k_1}) \ldots (Z_{k_j,i} - M_{k_j}) \right) \right] \cdot \end{aligned} \quad (13)$$

After plugging the previous expression into Equation 8 and taking expectation, we conclude that the AME is approximated by the following formula

$$\begin{aligned} AME_k &= \hat{\beta}_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon E[\hat{\delta}_i(\mathbf{Z}_i' \hat{\gamma})] \\ &= MEAI_k - \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \sum_{j=1}^{\infty} \\ &\quad \times \left[\frac{1}{j!} \sum_{k_1,\ldots,k_j} \left(\left. \frac{\partial^j \delta_i(\mathbf{Z}_i' \hat{\gamma})}{\partial Z_{k_1,i}, \ldots, \partial Z_{k_j i}} \right|_{\mathrm{M}} \cdot \Psi^j_{k_1,\ldots,k_j} \right) \right] \\ &= MEAI_k + B_k^1(\Psi^1, \Psi^2, \ldots) \end{aligned} \quad (14)$$

where $\Psi^j_{k_1,\ldots,k_j} = E[(Z_{k_1,i} - M_{k_1}) \ldots (Z_{k_j,i} - M_{k_j})]$ denotes the jth order joint moment about the means, while B_k^1 denotes the size of the first order approximation asymptotic bias as a function of the joint moments, Ψ^j, of the individual characteristics. Therefore by using the $\widehat{MEAI}_k$ to estimate the AME_k, one implicitly takes into account only the first order approximation while neglecting the higher orders, which ultimately leads to bias equal to $\hat{B}_k^1(\Psi^1, \Psi^2, \ldots)$. If instead one used an additional term of the Taylor polynomial, the *second order approximation of the AME* ($\widehat{SOAME}_k$) would substitute the $\widehat{MEAI}_k$. That would be given by the following formula:

$$\begin{aligned} \widehat{SOAME}_k = \widehat{MEAI}_k - \frac{1}{2} \hat{\gamma}_k \hat{\rho} \hat{\sigma}_\varepsilon \sum_{k2} \\ \times \sum_{k_2} \left(\left. \frac{\partial^2 \hat{\delta}_i(\mathbf{Z'}_i \hat{\gamma})}{\partial Z_{k_1,i} \partial Z_{k_2,i}} \right|_{\bar{\mathbf{Z}}} \cdot \widehat{\mathrm{Cov}}(Z_{k_1,i}, Z_{k_2,i}) \right) \end{aligned} \quad (15)$$

By using the second order approximation, which does not significantly increase the number of numerical operations since it only involves the elements of the entrywise product of the Hessian evaluated at $\bar{\mathbf{Z}}$ and the covariance matrix, one would substantially reduce[2] the bias of the estimates.

In the following section we empirically show that neglecting the bias could create misleading results that could significantly affect the policy implications of the model.

IV. Empirical Applications

We divide our applications into two parts: a study of earnings assimilation of immigrants in Sweden, where we with the use of real data illustrate the necessity of bias reduction in the estimation of marginal effects, and a Monte Carlo simulation where we examine the limiting properties of our approximation technique.

Earnings assilmation of immigrants in Sweden

The economic performance of immigrants is one of the major interests of policy makers in most highly immigrated Western countries. The question in such a study would typically be whether immigrants entered the host country with an earnings difference relative to natives and whether their earnings converge to those of the natives while years since migration (YSM) increase (Borjas, 1985, 1999; Longva and Raaum, 2003). Then, based on the answer, policies targeting to different individual characteristics of the immigrants are designed, in order to adjust the speed of assimilation closer to what is desired by the policy makers.

The data used in the present study comes from the registered nationally representative longitudinal individual data set of Sweden (LINDA), which comes in panel form and is rich in individual socioeconomic characteristics (Edin and Frederiksson, 2001). The principal data sources are income registers and population censuses. Family members are included in the sample only as long as they stay in the household. LINDA contains a sub-panel of about 20% of the foreign-born population. The working sample includes 3136 male individuals, aged 18–65 (1962 immigrants[3] and 1174 natives) followed for 11 years from 1990 to 2000.

Table 1 shows the mean characteristics of the sample. The earnings and the income from other sources are considerably higher among natives than among immigrants. Natives are more likely to be employed (0.82 vs. 0.57), are slightly older (38.4 vs. 37.1), but are also less likely to be married (0.39 vs. 0.43) and they have fewer children at home (0.44 vs. 0.48). They also acquire a higher level of education: 76% of natives are high school graduates, while the number is 71% among immigrants.

The immigrant arrival cohorts are classified into five year intervals except for the first and the last ones, which include the years before 1970 and the 1995–2000 period (six years), respectively. These two cohorts are slightly un-derrepresented in the sample (7 and 6% respectively). The immigrants are categorized according to their country of origin as follows: Nordic countries, USA, Western countries except USA (EU-15, Canada, Australia and New Zealand), Eastern Europe, Middle East, Asia, Africa and Latin America.

[2] The expected second order of magnitude is larger than the third one (Nguyen and Jordan, 2003).
[3] We define an immigrant as an individuals who was born abroad (first generation).

Table 1. Mean characteristics of immigrants and natives

Variables	Immigrants		Natives	
	Mean	SD	Mean	SD
Log earnings	8.5707	5.2519	10.7750	3.7428
Log nonlabour income	0.5656	1.9748	0.7746	2.3281
Employment	0.5713	0.4991	0.8221	0.4871
Age	0.3714	0.1103	0.3837	0.1127
Age squared	0.1501	0.0866	0.1599	0.0907
Big city (>250 000)	0.6347	0.4815	0.7349	0.4414
Number of children	0.4840	0.9875	0.4407	0.8959
Married/cohabiting	0.4344	0.4957	0.3891	0.4876
YSM	0.0794	0.0918	–	–
YSM squared	0.0147	0.0247	–	–
Education (highest level)				
Lower-secondary	0.2955	0.4852	0.2389	0.4911
Upper-secondary	0.4454	0.4970	0.4867	0.4998
University	0.2591	0.4381	0.2744	0.4462
Arrival cohort				
<1970	0.0669	0.2496	–	–
1970–1974	0.1176	0.3221	–	–
1975–1979	0.1574	0.3642	–	–
1980–1984	0.1372	0.3441	–	–
1984–1989	0.2237	0.4351	–	–
1990–1994	0.2335	0.4411	–	–
1995–2000	0.0637	0.1857	–	–
Geographical origin				
Nordic	0.1239	0.3609	–	–
W. Europe (incl. EU)	0.1188	0.2353	–	–
USA	0.1312	0.2485	–	–
Eastern Europe	0.1276	0.3337	–	–
Middle East	0.1434	0.3505	–	–
Asia	0.1245	0.3412	–	–
Africa	0.1250	0.3418	–	–
Latin America	0.1056	0.3097	–	–

Based on working indicators in the data, an employment dummy is defined that takes a value of 1 if the individual is employed and 0 otherwise. The earnings variable used in the study is obtained from the national tax registers and is measured in thousands of Swedish Kroner (SEK) per year, adjusted to 2000 prices.

The model specification for the immigrants is given by the following standard sample selection model:

$$
\begin{aligned}
Y_i^* &= \mathbf{X}_i'\beta + \phi AGE_i + \delta YSM_i + \sum_j \psi_j C_i^j + \sum_k \theta_k \Pi_i^k + \varepsilon_i \\
H_i^* &= \mathbf{Z}_i\gamma + u_i \\
H_i &= 1[H_i^* > 0] \\
Y_i &= Y_i^* \cdot H_i, \qquad (16)
\end{aligned}
$$

where i denotes each cross section, and Y^* is the natural logarithm of the latent earnings. The individual characteristics included in the $\mathbf{X}_i$ matrix are individual i's number of children, marital status, size of permanent residence, education, and geographical origin. The variables AGE and YSM denote the age and the years since migration respectively.[4] Finally C_i^j and Π_i^k are indicator variables for the j-th immigrant arrival cohort and the k-th year. C_i^j becomes 1 if the individual arrived at the j-th cohort and 0 otherwise. Similarly, Π_i^k takes the value 1 if the individual is observed in the k-th period, and the value 0 otherwise. The $\mathbf{Z}_i$ matrix includes the same characteristics plus the logarithm of nonlabour income.[5] The model specification for the natives does not differ from the one estimated for the immigrants, with the exception of the variables that are not applicable, e.g. years since immigration, arrival cohort and geographical origin.

The assimilation model given by (16) aims to identify the three important effects (aging, arrival cohort and period effect) on the earnings assimilation simultaneously. However, this model is not identified in any given cross section, since the calendar year in which the cross section is observed is the sum of YSM in the host country and the calendar year in which the individual immigrated. Thus the identification restriction imposed in the present study is that the period effect in the immigrant earnings equation is equal to that of the natives ($\Pi_i^I = \Pi_i^N, \forall i = 1, \ldots, 11$), which is a standard assumption in the assimilation literature (Borjas, 1985, 1999).

The estimation results and the bias analysis for the probit equation (first step) and the target equation (second step) are presented in respectively, along with the $\widehat{AME}$, the $\widehat{MEAI}$, the $\widehat{SOAME}$ and the first and second order bias ($\widehat{FOBIAS}$ and $\widehat{SOBIAS}$), which denote the difference between the consistent estimator $\widehat{AME}$ and its first ($\widehat{MEAI}$) and second order ($\widehat{SOAME}$) approximations respectively. For example, the $\widehat{AME}$ for the variable AGE for the immigrants is estimated to 0.1 53, while the corresponding $\widehat{MEAI}$ and $\widehat{SOAME}$ are equal to 0.235 and 0.175 respectively, which constituting a 73% improvement of the bias.

Taking a closer look at the first and second order bias estimates of the selection and the earnings equation (respectively), one can easily notice the rather significant improvement in all variables, not only in relative but also in absolute terms.

[4] The exact functional forms for age and years since migration are quadratic. The second order terms are omitted for notation simplicity purposes.

[5] The exclusion restriction adopted in this article is that the nonlabour income affects the probability of being employed but not the earnings.

Table 2. Estimates and analysis of bias for the employment equations

Variables	Est.	*AME*	*MEAI*	*SOAME*	*FO Bias*	*SO Bias*
Immigrants						
Constant	−1.3258	−0.3387	−0.5195	−0.3871	0.1808	0.0485
Log nonlabour income	−0.7741	−0.1977	−0.3033	−0.2260	0.1055	0.0283
Age	0.1259	0.1530	0.2347	0.1749	−0.0817	−0.0289
Age squared	−0.0016	–	–	–	–	–
Big city (>250 000)	0.1115	0.0285	0.0437	0.0326	−0.1520	−0.0041
Number of children	−0.0170	−0.0044	−0.0067	−0.0050	0.0023	0.0006
Married/cohabiting	0.3598	0.0919	0.1410	0.1051	−0.0490	−0.0132
YSM	0.0477	0.0122	0.0187	0.0139	−0.0065	−0.0017
YSM squared	−0.0001	–	–	–	–	–
Education (highest level)						
Upper-secondary	0.3657	0.0934	0.1433	0.1068	−0.0499	− 0.0134
University	0.5363	0.1370	0.2101	0.1566	−0.0731	− 0.0196
Arrival cohort						
1970–1974	−0.2306	−0.0589	−0.0904	0.0314	−0.0673	0.0084
1975–1979	−0.2826	−0.0722	−0.1107	−0.0825	0.0385	0.0103
1980–1984	−0.3285	−0.0839	−0.1287	−0.0959	0.0448	0.0120
1985–1989	−0.3510	−0.0897	−0.1375	−0.1025	0.0479	0.0128
1990–1994	−0.7965	−0.2035	−0.3121	−0.2326	0.1086	0.0291
1995–2000	−0.6630	−0.1694	−0.2598	−0.1936	0.0904	0.0242
Geographical origin						
Nordic	−0.8735	−0.2231	−0.3422	−0.2551	0.1191	0.0319
W. Europe (incl. EU)	−0.9631	−0.2461	−0.3774	−0.2813	0.1313	0.0352
USA	−0.3394	−0.3422	−0.5248	−0.3912	0.1826	0.0490
Eastern Europe	−0.3023	−0.3327	−0.5103	−0.3803	0.1776	0.0476
Middle East	−1.5686	−0.4007	−0.6146	−0.4581	0.2139	0.0573
Asia	−1.1450	−0.2925	−0.4486	−0.3344	0.1561	0.0419
Africa	−1.4546	−0.3716	−0.5699	−0.4248	0.1983	0.0532
Latin America	−1.1511	−0.2941	−0.4510	−0.3362	0.1569	0.0421
Natives						
Constant	−1.8781	−0.2753	−0.5145	−0.4719	0.2392	0.1966
Log nonlabour income	−0.8216	−0.1204	−0.2251	−0.2064	0.1046	0.0860
Age	0.1480	0.0016	0.0029	0.002741	−0.0014	−0.0011
Age squared	−0.0018	–	–	–	–	–
Big city	0.0801	0.0118	0.0220	0.0201	−0.0102	−0.0084
Number of children	0.0551	0.0080	0.0151	0.0139	−0.0070	−0.0058
Married/cohabiting	0.3974	0.0583	0.1089	0.0999	−0.0506	−0.0416
Education (highest level)						
Upper-secondary	0.3803	0.0557	0.1042	0.0956	−0.0484	−0.0398
University	0.4964	0.0728	0.1360	0.1247	−0.0632	−0.0520

Notes: The estimated average marginal effects (*AME*), marginal effects for the average individual (*MEAI*), the second order approximation of the average marginal effects (*SOAME*), and first (*FO Bias*) and second (*SO Bias*) order bias are presented in the table. The estimated SEs can be provided upon request.

This becomes even more worth mentioning since it is observed in the key variables. For instance, having a university degree improves the earnings of the immigrants by 0.340 log points, according to the $\widehat{AME}$. On the other hand, using the $\widehat{MEAI}$ yields an estimate equal to 0.370 log points. Finally, the $\widehat{SOAME}$ is equal to 0.348, which is substantially closer to the $\widehat{AME}$ (73% bias correction).

A really interesting result, though not surprising given the structure of the Taylor series, is that the percentage change in the bias level by shifting to the second order approximation remains constant across explanatory variables. Table A1 shows the size of the relative improvement when the second order approximation is used.

As we mentioned earlier, the hypothesis that one is usually willing to test in this specific type of study is *whether the earnings of the immigrants catch up with those of the natives with enough years spent in the host country, and if so how long this assimilation process takes*. Assume that the aging variables are defined as a function of time ($AGE(t)$ and $YSM(t)$). Then the relative earnings for immigrant i with respect to

Table 3. Estimates and analysis of bias for the earnings equations

Variables	Est.	*AME*	*MEAI*	*SOAME*	*FO Bias*	*SO Bias*
Immigrants						
Constant	11.5815	11.1524	11.0788	11.1330	0.0737	0.0195
Age	0.0290	0.0130	0.0132	0.0131	−0.0001	−0.00004
Age squared.	−0.0002	–	–	–	–	–
Big city (>250 000)	−0.0541	−0.0181	−0.0119	−0.0165	−0.0062	−0.0016
Number of children	−0.0117	−0.0172	−0.0181	−0.0174	0.0009	0.0002
Married/cohabiting	0.0217	0.1381	0.1581	0.1434	−0.0200	−0.0053
YSM	0.0075	0.0229	0.0256	0.0236	−0.0026	−0.0007
YSM squared	0.0003	–	–	–	–	–
Education (highest level)						
Upper-secondary	−0.0242	0.0941	0.1145	0.0995	−0.0203	−0.0054
University	0.1665	0.3401	0.3699	0.3479	−0.0298	−0.0079
Arrival cohort						
1970–1974	0.0966	0.0220	0.0092	0.0186	0.0128	0.0033
1975–1979	0.1712	0.0797	0.0640	0.0756	0.0157	0.0042
1980–1984	0.2659	0.1597	0.1414	0.1548	0.0183	0.0048
1985–1989	0.3291	0.2155	0.1960	0.2103	0.0195	0.0052
1990–1994	0.4727	0.2150	0.1707	0.2032	0.0443	0.0117
1995–2000	0.6263	0.4118	0.3750	0.4021	0.0368	0.0097
Geographical origin						
Nordic	−0.4172	−0.6998	−0.7484	−0.7127	0.0485	0.0128
W. Europe (incl. EU)	−0.3966	−0.7082	−0.7618	−0.7223	0.0535	0.0142
USA	−0.3288	−0.7622	−0.8367	−0.7819	0.0744	0.0197
Eastern Europe	−0.4382	−0.8596	−0.9320	−0.8788	0.0723	0.0191
Middle East	−0.5098	−1.0174	−1.1045	−1.0404	0.0872	0.0231
Asia	−0.4402	−0.8107	−0.8744	−0.8276	0.0636	0.0168
Africa	−0.4732	−0.9439	−1.0247	−0.9653	0.0808	0.0213
Latin America	−0.5268	−0.8993	−0.9633	−0.9162	0.0640	0.0169
Natives						
Constant	12.1808	11.3733	11.1341	11.3868	0.2392	−0.0135
Age	0.0043	0.0147	0.0159	0.0146	−0.0012	0.0001
Age squared	0.0080	–	–	–	–	–
Big city	−0.0708	−0.0363	−0.0261	−6.7524	−0.0102	0.0006
Number of children	−0.0445	−0.0208	−0.0138	−0.0212	−0.0070	0.0004
Married/cohabiting	0.0260	0.1969	0.2475	0.1941	−0.0506	0.0029
Education (highest level)						
Upper-secondary	−0.0106	0.1529	0.2014	0.1502	−0.0484	0.0027
University	0.2361	0.4496	0.5128	0.4460	−0.0632	0.0036

Note: See the note of Table 2.

native j, t years after migration, are given by the following equation:

$$\Delta Y_{i,j}(t) = E^I\left[Y_i|H_i = 1, AGE(t_0+t), YSM(t), \mathbf{X}_i, \mathbf{Z}_i\right] - E^N\left[Y_j|H_j = 1, AGE(t_0+t), \mathbf{X}_j, \mathbf{Z}_j\right] \quad (17)$$

where t_0 is the age at migration,[6] while E^I and E^N denote the conditional expectations of the assimilation model of the immigrants and the natives respectively. Evaluating $\Delta Y_{i,j}(t)$ at $t=0$ yields the initial earnings difference, otherwise called entry effect upon arrival.

Then the estimated marginal rate of assimilation ($\widehat{MRA}$), which shows the rate of earnings convergence between the i-th immigrant and the j-th native at time t (Barth *et al.*, 2004), is given by the following equation:

$$\widehat{MRA}_{i,j}(t) = \frac{\partial E_i^I}{\partial t} - \frac{\partial E_j^I}{\partial t} \quad (18)$$

or in terms of marginal effects:

$$\widehat{MRA}_{i,j}(t) = \widehat{ME}^I_{AGE,i}(t) + \widehat{ME}^I_{YSM,i}(t) - \widehat{ME}^N_{AGE,j}(t) \quad (19)$$

[6] The entry age in the present study is assumed to be constant across immigrants and equal to 20.

We thus reach a point where the marginal effects are in question again. Given the fact that we are interested in the average total years of assimilation ($\widehat{ATYA}$), one should estimate the average marginal rate of assimilation ($\widehat{AMRA}$). Namely,

$$\begin{aligned}\widehat{AMRA}(t) &= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{I}\frac{1}{J} \\ &\quad \times \left(\widehat{ME}^{I}_{AGE,i}(t) + \widehat{ME}^{I}_{YSM,i}(t) - \widehat{ME}_{AGE^{N},j}(t)\right) \\ &= \frac{1}{I}\sum_{i=1}^{I}\widehat{ME}^{I}_{AGE,i}(t) + \frac{1}{I}\sum_{i=1}^{I}\widehat{ME}^{I}_{YSM,,i}(t) \\ &\quad - \frac{1}{J}\sum_{j=1}^{J}\widehat{ME}^{N}_{AGE,j}(t) \\ &= \widehat{AME}_{AGE}(t) + \widehat{AME}^{I}_{YSM}(t) - \widehat{AME}^{N}_{AGE}(t)\end{aligned} \tag{20}$$

where I and J denote the total number of immigrants and natives respectively. One can similarly calculate the estimators for the marginal rate of assimilation for the average individual ($\widehat{MRAAI}$) and the second order approximation of the average marginal rate of assimilation ($\widehat{SOAMRA}$), by substituting the corresponding marginal effects in Equation 20.

Then the estimator of the average total years of assimilation ($\widehat{ATYA}$) is the upper limit that equates the following integral with the average initial earnings difference:

$$\int_{0}^{\widehat{ATYA}} \widehat{AMRA}(t)\,\mathrm{d}t = \Delta Y(0) \tag{21}$$

Table A2 shows the estimation results. The $\widehat{ATYA}$ is reported in the first column for each group of immigrants. According to this estimator, the earnings of the immigrants from for example Africa catch up to the level of the natives *on average* 25.3 years after arrival. The second column of the table reports total years of assimilation for the average immigrant ($\widehat{TYAAI}$). The corresponding estimate for the *average* African immigrant is 23.6 years, which is 1.7 years shorter than the ($\widehat{ATYA}$). Finally, by using the method we propose in the present article, the second order approximation of the average total years of assimilation ($\widehat{SOATYA}$) yields an estimate of 24.4 years, which is 54% closer to the targeted result.

Monte Carlo simulation

As we have already discussed, the bias that emerges when using the $\widehat{MEAI}$ as a point estimator of the AME is not a consequence of a small sample, which would disappear in the limit. Regardless of the sample size, the second order approximation leads to bias reduction compared to the first one. The purpose of this section is to provide empirical evidence for the size of the bias reduction through a Monte Carlo experiment.

Assume a standard sample selection model of the form of Equation 1, with $\mathbf{X}_i$ being a singleton and $\mathbf{Z}_i = (Z_{1,i}, Z_{2,i})$ coming from the bivariate normal distribution with mean $\mu_i = (\mu_1, \mu_2)$ and covariance matrix Σ. Assume also the following parameter values: $\beta = 1$, $\gamma = (3, -2)$, $\sigma_\varepsilon = 0.5$, $\sigma_u = 1$, $\rho = -0.8$, $\mu = (0.5, 1.5)$ and $\Sigma = [\begin{smallmatrix}0.5 & -0.1\\ -0.1 & 1\end{smallmatrix}]$. By using pseudo-random numbers, we then repeatedly evaluate the first and the second order bias, while increasing the sample size in steps of 100 observations. The results are presented in Table A3.

Figure 1 illustrates the same point as Table A3, namely that it becomes clear that the bias that emerges when using the $\widehat{MEAI}$, is corrected to a rather large extent, without a corresponding

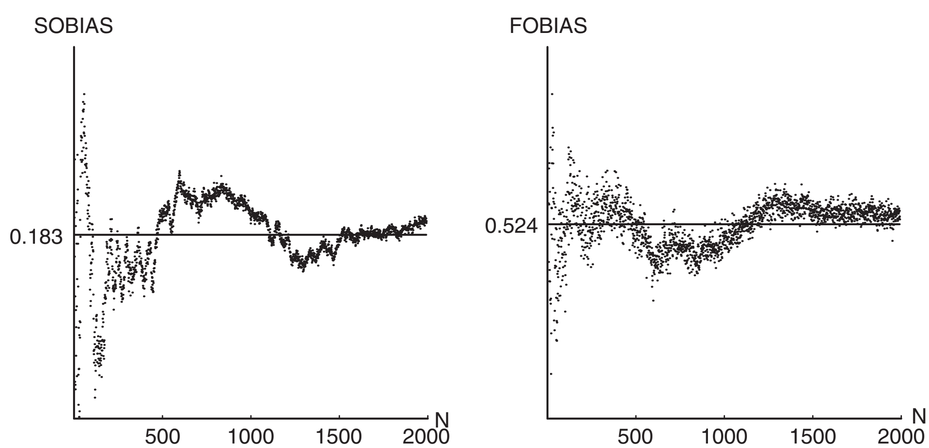

Fig. 1. First and second order bias in Monte Carlo experiment

computational cost. Notice that bias reduction is observed not only for small samples, but also asymptotically.

V. Concluding Discussion

In this article we discuss the differences between two point estimators of the marginal effect of an explanatory variable on the population, in a sample selection model estimated by Heckman's two step procedure. We show that contrary to a rather widespread perception that neglects any differences between them, the AME is significantly different from the marginal effect of the average individual, even asymptotically. Thus, it should be clear that there is not only a quantitative distinction but also a conceptual one between these measures. Given that the usual aim is to extract information about the average effects on the population, a clear bias would emerge if using the marginal effect of the sample average individual. Hence, we suggest an approximation method based on the Taylor expansion, which should correct the bias to a rather remarkable extent, while increasing the number of computational operations relatively little. Such an example is presented in the article, along with a Monte Carlo experiment, both supporting the previous argument. Before closing, we would like to make clear that we do not argue in favour of the AME and against the marginal effect of the average individual. Instead, our aim is to stress that once the AME has been chosen as an informative tool for policy making, the sample marginal effect of the average individual provides inconsistent estimations which can be corrected to a large extent by the proposed method.

Acknowledgements

We are indebted to Lennart Flood, Marcela Ibanez, Florin Maican and Kerem Tezic for their benefitting comments. We would also like to specially thank an anonymous referee for his valuable suggestions. All mistakes and misprints are exclusively ours.

References

Barth, E., Bratsberg, B., Raaum, O. (2004) Identifying earnings assimilation of immigrants under changing macroeconomic conditions, *Scandinavian Journal of Economics*, **106**, 1–22.

Borjas, G. J. (1985) Assimilation, changes in cohort quality, and the earnings of immigrants, *Journal of Labor Economics*, **3**, 463–89.

Borjas, G. J. (1999) *The economic analysis of immigration Handbook of Labor Economics*, Vol. 3A (Eds) O. Ashenfelter and D. Card, pp. 1697–760.

Chiswick, B. R. (1978) The effect of Americanization on the earnings of foreign-born men, *Journal of Political Economy*, **86**, 897–921.

Edin, P. A. and Frederiksson, P. (2001) LINDA – Longitudinal individual data for Sweden, Working Paper, Uppsala University, Department of Economics 2001:6.

Greene, W. H. (2003) *Econometric Analysis*, 5th edn, Prentice Hall, Saddle River.

Heckman, J. J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models, *Annals of Economic and Social Measurement*, **5**, 475–92.

Heckman, J. J. (1979) Sample selection bias as a specification error, *Econometrica*, **47**, 153–62.

Hoffmann, R. and Kassouf, A. L. (2005) Deriving conditional and unconditional marginal effects in log earnings equation estimated by Heckman's procedure, *Applied Economics*, **37**, 1303–11.

Longva, P. and Raaum, O. (2003) Earnings assimilation of immigrants in Norway-a reappraisal, *Journal of Population Economics*, **16**, 177–93.

Nguyen, X. and Jordan, M. I. (2003) On the concentration of expectation and approximate inference in layered networks, *Advances in Neural Information Processing Systems (NIPS)*, 16.

Saha, A., Capps Jr, O. and Byrne, P. J. (1997a) Calculating marginal effects in models for zero expenditures in household budgets using Heckman-type correction, *Applied Economics*, **29**, 1311–6.

Saha, A., Capps Jr, O. and Byrne, P. J. (1997b) Calculating marginal effects in dichotomous-continuous models, *Applied Economics Letters*, **4**, 181–5.

Verlinda, J. A. (2006) A comparison of two common approaches for estimating marginal effects in binary choice models, *Applied Economics Letters*, **13**, 77–80.

Appendix

Table A1. Relative reduction of the bias

	Immigrants	Natives
Selection equation	0.714	0.143
Earnings equation	0.943	0.735

Table A2. Estimates and analysis of bias for the assimilation period

Variables	Earn. diff.	*ATYA*	*TYAAI*	*SOATYA*	*FO Bias*	*SO Bias*
Nordic	0.2916	13.6973	12.7850	13.1966	0.9123	0.5006
W. Europe (incl. EU)	0.1851	8.6961	8.1169	8.3782	0.5792	0.3178
USA	0.1895	8.9012	8.3083	8.5758	0.5929	0.3253
Eastern Europe	0.3285	15.4322	14.4043	14.8682	1.0279	0.5641
Middle East	0.5099	23.9514	22.3561	23.0760	1.5953	0.8754
Asia	0.4449	20.8989	19.5069	20.1351	1.3920	0.7639
Africa	0.5392	25.3264	23.6395	24.4007	1.6869	0.9256
Latin America	0.4047	19.0115	17.7452	18.3166	1.2663	0.6949
Total	0.3617	16.9894	15.8578	16.3684	1.1316	0.6210

Notes: The initial earnings difference, the estimated average total years of assimilation (*ATYA*), total years of assimilation for the average immigrant (*TYAAI*), the second order approximation of the average total years of assimilation (*SOATYA*) and first (*FO Bias*) and second (*SO Bias*) order bias are presented in the table. The estimated SEs can be provided upon request.

Table A3. Bias convergence in Monte Carlo simulation

Number of obs.	*AME*	*MEAI*	*SOAME*	*FO Bias*	*SO Bias*	Rel. improv.
1000	1.4034	1.0060	1.2033	0.3974	0.2001	0.4965
10 000	1.5300	1.0100	1.3900	0.5160	0.1400	0.7308
50 000	1.5303	1.0 080	1.3392	0.5222	0.1910	0.6342
100 000	1.5343	1.0084	1.3500	0.5259	0.1843	0.6496
250 000	1.5321	1.0082	1.3436	0.5239	0.1886	0.6401
500 000	1.5338	1.0083	1.3488	0.5255	0.1850	0.6479

Note: See the note of Table 2.