

# Consequences of Specification Error for Distributional Analysis with an Application to Intergenerational Mobility.\*

Donal O'Neill<sup>†</sup>, Olive Sweetman<sup>‡</sup> and Dirk Van de gaer<sup>§</sup>

December 19, 2001

## Abstract

We analyze the consequences of three types of specification error for cumulative conditional distribution functions  $F(y|a)$ : measurement error in  $y$ , in  $a$  and omitted conditioning variables. The paper uses exact results to obtain conditions under which the effect of the misspecification on the computed cumulative distribution function can be signed. The effects are shown to depend on both the curvature of the true distribution and the properties of the error distribution. We illustrate our findings using a model of intergenerational mobility.

Keywords: Specification error, mobility, conditional distributions.

JEL Classification: C14

---

\*We thank Ramses Abul Naga and Christian Schluter, seminar participants at NUI Maynooth, the European Society of Population Economics meeting (Athens, June 2001), the North American Econometric Society meetings (University of Maryland, June 2001), the Econometric Society European Meeting (Lausanne, August 2001) and the ESF funded conference on Income Mobility (Bristol December 2001) for helpful comments on an earlier version of this paper.

<sup>†</sup>N.U.I. Maynooth, Maynooth, Co. Kildare, Ireland

<sup>‡</sup>N.U.I. Maynooth, Maynooth, Co. Kildare, Ireland

<sup>§</sup>Ghent University, F.E.B. Vakgroep Sociale Economie, Hoveniersberg 24, B-9000 Gent, Belgium

# 1 Introduction

A number of recent studies have examined the relationship between variables by estimating the entire joint distribution rather than restricting attention to conditional means. This approach has been used to look at changes in income distributions over time using both micro data (DiNardo et al (1996) and Burkhauser et al (1999)) and macro data (Quah (1997)). The approach has also been used to examine mobility patterns for a given individual over time (Trede (1998a and 1998b)) and mobility across generations (Corak and Heisz (1999) and O'Neill, Sweetman and Van de gaer (2000)). However, despite the increasing popularity of distributional analysis, there appears to be little research on the consequences of measurement error or other forms of misspecification for this approach. Exceptions include Chesher (1991) who studies the effect of measurement error on probability distributions and Magnac and Visser (1999) who control for measurement error when estimating transition probabilities.

In this paper we examine the properties of conditional distribution functions in the presence of several types of misspecification. We derive the properties of the cumulative conditional distribution function in the case of measurement error in the dependent variable, measurement error in the conditioning variables and omitted conditioning variables. Our focus on the conditional cumulative distribution,  $F(y | a)$ , is motivated by the fact that this distribution provides the natural starting point to investigate the direction and intensity of the effect of the variable  $a$  on the variable  $y$ . The results we present on measurement error have the character of dominance results: if the distribution of measurement error satisfies certain properties then we are able to sign the effect of specification error for a wide class of conditional CDF's.

In the literature on measurement error it is often assumed that the mean error is zero. Often the stronger assumption that the distribution is symmetric around zero is used. In some important contexts, these are strong assumptions. For example, it is well known that people under declare their incomes, such that the expected error is negative. We derive results that are applicable to such cases, thereby throwing light on the restrictiveness of the traditional assumptions about measurement error.

We illustrate our findings using a model of intergenerational mobility. There have been many papers which have used micro data to estimate the

intergenerational elasticity of incomes<sup>1</sup>. Many of these studies use linear regression techniques. Usually one is interested in the elasticity of lifetime incomes of children with respect to their parents' lifetime income. However it is often difficult to obtain accurate measures of lifetime incomes. As a result both the parent's and child's lifetime incomes tend to be measured with error. Several authors (e.g., Atkinson, Maynard and Trinder (1983, p.84-87), Zimmerman (1992) and Solon(1992)) discuss the likely consequences of measurement error in fathers' incomes. Measurement error in the son's income does not seem to be a cause for concern in this literature presumably because the resulting OLS estimator is still unbiased.<sup>2</sup> It is also well known that omitting endowments, such as ability, may bias the estimate of the intergenerational elasticity of incomes.<sup>3</sup>

There have also been some studies that have used transition matrices to examine intergenerational mobility patterns (see, e.g., Atkinson et al. (1983) and Zimmerman (1992)). Some models of intergenerational mobility predict that regression to the mean may differ depending on the level of father's income (Becker and Tomes (1986)). Using transition matrices allows the researcher to examine mobility at different points of the distribution. Recent studies (Corak and Heisz (1999) and O'Neill, Sweetman and Van de gaer (2000)) have extended this approach by estimating the entire joint distribution of father's and son's earnings. However, little is known about the consequences of specification error for these approaches.

The results presented in Section 2 of the paper show the effect of specification error on the cumulative conditional distribution function. The effects are shown to depend on both the curvature of the true distribution and the properties of the error distribution. We then present a simple intergenerational model to illustrate our findings. Section 4 examines the consequences of specification error for the measurement of intergenerational mobility. Section 5 concludes the paper.

---

<sup>1</sup>For a review of this literature see Solon (1999). Abul Naga (2001) provides a critical review of the standard procedures used in the literature.

<sup>2</sup>For a more detailed discussion of measurement error issues see Fuller (1987). Griliches and Ringstad (1970) and Hausman, Newey and Powell (1995) focus on the effects of measurement error in non-linear models.

<sup>3</sup>For a recent discussion of this see Han and Mulligan (2000).

## 2 Properties of conditional cumulative distributions

Let  $Y, A$  and  $B$  be three random variables. The joint distribution of these random variables can be described by the cumulative distribution  $F_{y^*, a^*, b^*}(y, a, b) : ([\underline{y}, \bar{y}] \times [\underline{a}, \bar{a}] \times [\underline{b}, \bar{b}]) \rightarrow [0, 1]$ . Initially, we focus on the computation of the conditional cumulative distribution  $F_{y^*, a^*}(y | a) : ([\underline{y}, \bar{y}] \times [\underline{a}, \bar{a}]) \rightarrow [0, 1]$ . In particular we concentrate on three types of misspecification. Firstly  $y^*$  might be unobservable, and we may have to use observations on a related variable,  $y$ , to compute a cumulative density. Alternatively,  $a^*$  might not be observable and we may be forced to condition on a related observable variable,  $a$ . These problems are analyzed in subsections one and two below. Finally, we might be interested in the distribution of  $y^*$ , conditional upon two variables,  $a^*$  and  $b^*$ . In certain circumstances  $b^*$  may be unobservable. The third subsection analyses the position of the conditional cumulative distribution  $F_{y^*, a^*}(y | a)$ , relatively to the distribution  $F_{y^*, a^*, b^*}(y | a, b) : ([\underline{y}, \bar{y}] \times [\underline{a}, \bar{a}] \times [\underline{b}, \bar{b}]) \rightarrow [0, 1]$ . We discuss the theorems and corollaries in the body of the text. Their formal proofs are gathered in appendix one.

### 2.1 Definitions and Notation

In the next two subsections we want to derive conditions on the distribution of measurement error that allow us to determine the relationship between the cumulative conditional distribution of the variable of interest and the observed cumulative distribution. To do this we introduce the following definitions.

Let  $f_Z(\mathbf{Z}) : \mathfrak{R}^n \rightarrow \mathfrak{R}^+$  be a function in  $n$  variables.

#### Definition 1

(a)  $f_Z(\mathbf{Z})$  is pointwise left dominant (PLD) in  $z_i$  around  $z_i^0$  over  $[a, b]$  if and only if  $\forall r \in [z_i^0 - \max\{|a|, |b|\}, z_i^0 + \max\{|a|, |b|\}] :$

$$f_Z(z_1, \dots, z_i^0 - r, \dots, z_n) \geq f_Z(z_1, \dots, z_i^0 + r, \dots, z_n)$$

(b)  $f_Z(\mathbf{Z})$  is pointwise right dominant (PRD) in  $z_i$  around  $z_i^0$  over  $[a, b]$  if and only if  $\forall r \in [z_i^0 - \max\{|a|, |b|\}, z_i^0 + \max\{|a|, |b|\}] :$

$$f_Z(z_1, \dots, z_i^0 - r, \dots, z_n) \leq f_Z(z_1, \dots, z_i^0 + r, \dots, z_n)$$

(c)  $f_Z(\mathbf{Z})$  is pointwise symmetric (PS) in  $z_i$  around  $z_i^0$  over  $[a, b]$  if and only if  $f_Z(\mathbf{Z})$  is both PLD and PRD in  $z_i$  around  $z_i^0$  over  $[a, b]$ .

We will work with one or two dimensional functions most of the time. Let  $\Psi^1$  ( $\Psi^2$ ) denote the set of all one (two) dimensional positively valued functions. Often, we use the concepts of PLD or PRD in the context of density functions. In that context, PLD (PRD) in  $z_i$  around  $z_i^0$  over  $[a, b]$  means that the density function has more probability mass to the left (right) of  $z_i^0$  than to the right (left). (c) becomes the standard definition of symmetry of the distribution of  $z$  in dimension  $z_i$  around  $z_i^0$  over  $[a, b]$ . We can also define absolute pointwise dominance as follows:

### Definition 2

(a)  $f_Z(\mathbf{Z})$  is absolute pointwise left dominant (APLD) in  $z_i$  around  $z_i^0$  over  $[a, b]$  if and only if

$|f_Z(\mathbf{Z}) - f_Z(z_1, \dots, z_i^0, \dots, z_n)|$  is PLD in  $z_i$  around  $z_i^0$  over  $[a, b]$ .

(b)  $f_Z(\mathbf{Z})$  is absolute pointwise right dominant (APRD) in  $z_i$  around  $z_i^0$  over  $[a, b]$  if and only if

$|f_Z(\mathbf{Z}) - f_Z(z_1, \dots, z_i^0, \dots, z_n)|$  is PRD in  $z_i$  around  $z_i^0$  over  $[a, b]$ .

(c)  $f_Z(\mathbf{Z})$  is absolute pointwise symmetric (APS) in  $z_i$  around  $z_i^0$  over  $[\underline{z}_i, \bar{z}_i]$  if and only if

$|f_Z(\mathbf{Z}) - f_Z(z_1, \dots, z_i^0, \dots, z_n)|$  is PS in  $z_i$  around  $z_i^0$  over  $[a, b]$ .

Absolute pointwise dominance turns out to be a crucial property of cumulative conditional distribution functions when signing the effects of measurement error and is closely related to the curvature of the distribution. In particular functions that are concave (convex) in  $z_i$  are also PLD (PRD) in  $z_i$  although the reverse need not hold.

## 2.2 Measurement error in the $y^*$ -variable

In this section we are interested in computing the distribution of  $y^*$  conditional on  $a^*$  which we denote by  $F_{y^*, a^*}(y | a^0)$ . However, suppose that for some reason  $y^*$  cannot be measured accurately. Instead we observe  $y$  which is defined as  $y = y^* + u$ , where  $u$  is measurement error, with support  $[\underline{u}, \bar{u}]$ . We assume that  $u$  is not correlated with  $(y^*, a^*)$ . Our data therefore allow

us to compute  $F_{y,a^*}(y | a^0)$ . We establish the following theorem relating  $F_{y^*,a^*}(y^0 | a^0)$  to  $F_{y,a^*}(y^0 | a^0)$ :<sup>4</sup>

**Theorem 3** *Measurement error in the  $y^*$ -variable*

$$\Delta^y(y^0 | a^0) = F_{y,a^*}(y^0 | a^0) - F_{y^*,a^*}(y^0 | a^0) \quad (1)$$

$$= \int_{\underline{u}}^{\bar{u}} (F_{y^*,a^*}(y^0 - u | a^0) - F_{y^*,a^*}(y^0 | a^0)) f_u(u) du \quad (2)$$

Theorem 3 can be used to establish a number of results. Firstly at points below but close to  $\underline{y}$  we will overestimate the true CDF. At points above but sufficiently close to  $\bar{y}$ , we will underestimate the true CDF. The theorem yields another straightforward result. Suppose that  $\underline{u} < \bar{u} \leq 0$ , as would be the case if the observed value  $y$  was a lower bound for the true value  $y^*$ . In that case  $F_{y^*,a^*}(y^0 - u | a^0) > F_{y^*,a^*}(y^0 | a^0)$  and  $\Delta^y > 0$ . Use of the proxy for  $y^*$  results in an overestimation of the conditional cumulative density since with  $u \leq 0$ ,  $y \leq y^*$ . The opposite conclusion follows when  $0 \leq \underline{u} < \bar{u}$ .

Expression (2) shows that  $\Delta^y(y^0 | a^0)$  is a weighted average of deviations of  $F_{y^*,a^*}(y^0 - u | a^0)$  from  $F_{y^*,a^*}(y^0 | a^0)$  with the weights given by the density of  $u$ . To obtain an unambiguous sign for (2), restrictions have to be imposed on the distribution functions. Let  $\mathcal{F}$  be the set of all conditional distributions  $F(y | a)$ . We will obtain results for the families of CDF's defined below.

**Definition 4**

$$\begin{aligned} D_y^{AL} &= \{F(y | a^0) \in \mathcal{F} : F(y | a) \text{ is APLD in } y \text{ around } y^0 \text{ over } [y^0 - \bar{u}, y^0 - \underline{u}]\}. \\ D_y^{AR} &= \{F(y | a^0) \in \mathcal{F} : F(y | a) \text{ is APRD in } y \text{ around } y^0 \text{ over } [y^0 - \bar{u}, y^0 - \underline{u}]\}. \\ D_a^{AL} &= \{F(y^0 | a) \in \mathcal{F} : F(y | a) \text{ is APLD in } a \text{ around } a^0 \text{ over } [a^0 - \bar{u}, a^0 - \underline{u}]\}. \\ D_a^{AR} &= \{F(y^0 | a) \in \mathcal{F} : F(y | a) \text{ is APRD in } a \text{ around } a^0 \text{ over } [a^0 - \bar{u}, a^0 - \underline{u}]\}. \end{aligned}$$

As noted earlier  $D_y^{AL}$  contains all cumulative distribution functions that are concave in its first argument ( $y$ ) over  $[y^0 - \bar{u}, y^0 - \underline{u}]$ , while  $D_y^{AR}$  contains all distribution functions that are convex in  $y$  over this interval. Both sets

---

<sup>4</sup>The assumption that the error term is additive is less restrictive than it appears. In the case of mismeasured incomes it is often assumed that the error term enters multiplicatively (see, e.g., Chesher and Schluter (2001)). In this case our theorems would apply to the log transformed variables. Furthermore since  $F(y^0) > F(y^1)$  iff  $F(\log(y^0)) > F(\log(y^1))$  we can be sure that our qualitative results on the sign of  $F_{y,a^*}(y_0|a_0) - F_{y^*,a^*}(y_0|a_0)$  apply to the original untransformed variable with multiplicative errors.

of cumulative conditional distributions also contain conditional distribution functions that are not concave or convex, however. Cumulative distributions that are concave in the conditioning variable over  $[a^0 - \bar{u}, a^0 - \underline{u}]$  belong to  $D_a^{AL}$ , those that are convex over that interval belong to  $D_a^{AR}$ . Given these definitions we can establish the following result.

**Corollary 5**

- (a) If  $f_u(u) \in \{f_z(z) \in \Psi^1 \mid f_z(z) \text{ is PRD around } 0\}$  and  $F_{y^*,a^*}(y \mid a^0) \in D_y^{AL}$  then  $\Delta^y(y^0 \mid a^0) \leq 0$ .
- (b) If  $f_u(u) \in \{f_z(z) \in \Psi^1 \mid f_z(z) \text{ is PLD around } 0\}$  and  $F_{y^*,a^*}(y \mid a^0) \in D_y^{AR}$  then  $\Delta^y(y^0 \mid a^0) \geq 0$ .

Given that distributions that are symmetric around 0 are both PLD and PRD around 0 we can also establish the following theorem.

**Corollary 6**

- (a) If  $f_u(u)$  is symmetric around 0 and if  $F_{y^*,a^*}(y \mid a^0) \in D_y^{AL}$  then  $\Delta^y(y^0 \mid a^0) \leq 0$ .
- (b) If  $f_u(u)$  is symmetric around 0 and if  $F_{y^*,a^*}(y \mid a^0) \in D_y^{AR}$  then  $\Delta^y(y^0 \mid a^0) \geq 0$ .

This implies for instance that if  $u$  is symmetrically distributed with mean zero and  $F_{y^*,a^*}(y \mid a^0)$  is concave then we will underestimate the true CDF at every point.

Cumulative CDF's that are APS in  $y$  around  $y^0$  over  $[y^0 - \bar{u}, y^0 + \underline{u}]$  belong to both  $D_y^{AL}$  and  $D_y^{AR}$ . It follows immediately from corollary 6 that for such distributions  $\Delta^y(y^0 \mid a^0) = 0$ . In practice it is often assumed that  $F_{y^*,a^*}(y \mid a^0)$  is APS in  $y$  around  $E(y^* \mid a^0)$  and that  $f_u(u)$  is symmetric. In this case the following applies.

**Corollary 7** *If  $f_u(u)$  is symmetric around 0 and  $F_{y^*,a^*}(y \mid a^0)$  is APS in  $y$  around  $E(y^* \mid a^0)$  over  $[E(y^* \mid a^0) - \bar{u}, E(y^* \mid a^0) - \underline{u}]$ , then  $\Delta^y(E\{y^* \mid a^0\} \mid a^0) = 0$ .*

It is important to realize that whether  $F_{y^*,a^*}(y \mid a^0)$  belongs to  $D_y^{AL}$ ,  $D_y^{AR}$ , to both or to neither will, in general, depend on the value for  $y^0$ . Suppose, for instance that  $F_{y^*,a^*}(y \mid a^0)$  the CDF is of a uniform distribution

over  $[\underline{y}, \bar{y}]$ . Then, provided the support of  $u$  is not too big, this CDF will be in  $D_y^{AL}$  close to  $\bar{y}$ , but in  $D_y^{AR}$  close to  $\underline{y}$ .

The results in Corollary 5, Corollary 6 and Corollary 7 are exact results that rely on the properties of the error distribution. Chesher (1991) establishes a result similar to corollary 6 for concave and convex cumulative distributions where the measurement error has mean zero. As was stated earlier our sets of cumulative CDF's,  $D_y^{AL}$  and  $D_y^{AR}$  are larger than the set of concave or convex cumulative CDF's, respectively, so that our results are more general. Chesher's results are based on second order approximations that depend only on the curvature of the true function and the variance of the error distribution. The importance of the form of the error distribution in our results stems from the fact that they are exact results. This insight tends to get lost if one takes a second order approximation to the distribution. For instance even if the error has mean zero, if the distribution of measurement error is skewed to the right it can never be PRD around zero. Corollary 6 and corollary 5 (a) cannot be applied. Only Corollary 5 (b) can still hold true. If it skewed to the left then the opposite is true and only 5 (a) can hold. To see this consider the following example. Suppose the true distribution of  $y$  is standard normal. Consider two alternative forms of measurement error. The first is distributed normally with mean 0 and variance 2, the second follows a Chi-squared distribution with one degree of freedom, adjusted so as to have mean zero. Both these distributions have the same first two moments but the latter is skewed to the right. The symmetry of the normal distribution means that corollary 6 can be used to determine the relative bias. However the discussion above suggests that the skewness associated with the Chi-Squared distribution may cause problems in trying to sign the bias, especially in regions where the true underlying function is Absolute Pointwise Left Dominant. To examine this we randomly drew 1000 observations from the three relevant distributions and estimated the misspecified models. Figure 1 compares the misspecified models with the true model (standard normal). The results are as predicted. With the normal measurement error we overestimate in the region where the true CDF is convex and underestimate in the region where it is concave. This is in line with Chesher (1991). Note however that when the distribution of the measurement error is skewed to the right we cannot determine the sign of the



bias in the region where the true distribution is concave<sup>5</sup>. This is despite the fact that the true distribution is the same in both cases, as are the first two moments of the error distribution. The importance of the functional form of the error distribution is apparent in our global results but can get lost in approximations.

Similar problems can arise in situations where people are more likely to underreport (overreport) their income. Such a pattern of errors implies that the mean, median and/or mode of  $u$  will be less (greater) than zero. As a result  $f_u(u)$  cannot be pointwise right (left) dominant around zero and therefore only a subset of corollaries 5 and 6 can apply.

Sometimes it may be possible to establish the properties of the distribution of measurement error on the basis of auxillary data. It may then be possible examine the class of distributions to which the error distribution belongs which in conjunction with the results established above would aid in determining the sign of the bias.

A local result can be obtained under the following condition.

**Condition 8**  $F_{y^*,a^*}(y | a^0)$  is differentiable up to order  $n$  with respect to  $y$  over the domain  $[y^0 + \underline{u}, y^0 + \bar{u}]$ .

Define

$$F_{y^*,a^*}^{(k,y)}(y | a^0) \equiv \frac{\partial^k F_{y^*,a^*}(y | a^0)}{(\partial y)^k} \text{ and } m_u^k \equiv \int_{\underline{u}}^{\bar{u}} u^k f_u(u) du \quad \forall k \in \{1, 2, \dots, n\}$$

$F_{y^*,a^*}^{(k,y)}(y | a^0)$  is the  $k$ -th order partial derivative of the cumulative conditional distribution of  $y$  with respect to  $y$  and  $m_u^k$  is the  $k$ -th uncentered moment of the distribution of  $u$ . Then we can show that

**Corollary 9** If  $F_{y^*,a^*}(y | a^0)$  is differentiable up to order  $n$  with respect to  $y$  over the domain  $[y^0 + \underline{u}, y^0 + \bar{u}]$ , then

$$\Delta^y(y^0 | a^0) \simeq \sum_{k=1}^n (-1)^k \frac{1}{k!} \left| F_{y^*,a^*}^{(k,y)}(y | a^0) \right|_{y=y_0} m_u^k$$

From the corollary it follows that the error in the computed cumulative conditional density depends on the moments of the distribution of  $u$  and

---

<sup>5</sup>The  $\chi^2$  distribution is PLD around 0 over a sizeable interval, though not over  $[-\infty, +\infty]$ . The contribution to the sign of  $\Delta^y$  over this sizeable interval dominates, however. This explains why to the left of  $m_y$ , where the true CFDF is convex,  $\Delta^y > 0$ .

the properties of  $F_{y^*,a^*}(y | a^0)$ . Suppose that  $m_u^1 = 0$  and that the density  $f_{y^*,a^*}(y | a^0)$  is linear in  $y$  over the interval  $[y^0 + \underline{u}, y^0 + \bar{u}]$ .  $f_{y^*,a^*}(y | a^0)$  might be the uniform or a triangular density over this interval. In that case, the measurement error has no effect on the computed cumulative conditional distribution. If a second order approximation to  $F_{y^*,a^*}(y^0 - u | a^0)$  is sufficiently close only the first two moments matter of the distribution of  $u$ , not the higher order moments. Generally speaking, all moments of the distribution of  $u$  are important to establish whether  $\Delta^y(y^0 | a^0)$  is positive or negative, however.

Corollary 9 shows that the effect of the first moment being less than zero is to push  $\Delta^y(y^0 | a^0)$  upwards. The result is that overestimation of the true CDF becomes more likely.

### 2.3 Measurement error in the $a^*$ -variable

As before, we are interested in  $F_{y^*,a^*}(y | a^0)$ . The problem now is that we cannot observe  $a^*$ . Instead we observe  $a$ . We know that  $a = a^* + u$  and that  $u$  is uncorrelated with  $(y^*, a^*)$ . The following theorem establishes the relationship between  $F_{y^*,a}(y^0 | a^0)$  and  $F_{y^*,a^*}(y^0 | a^0)$ .

**Theorem 10** *Measurement error in the  $a^*$ -variable*

$$\begin{aligned} \Delta^a(y^0 | a^0) &= F_{y^*,a}(y^0 | a^0) - F_{y^*,a^*}(y^0 | a^0) \\ &= \frac{1}{\int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du} \int_{\underline{u}}^{\bar{u}} (F_{y^*,a^*}(y^0 | a^0 - u) - F_{y^*,a^*}(y^0 | a^0)) f_{a^*}(a^0 - u) f_u(u) du \end{aligned} \quad (3)$$

If we evaluate  $\Delta^a(y^0 | a^0)$  at points outside the support of  $a^*$ , we obtain a result that is similar to the one in the previous section. At points below  $\underline{a}$  but sufficiently close to  $\underline{a}$ , we overestimate the true CDF, while at points above  $\bar{a}$ , but sufficiently close to  $\bar{a}$  we underestimate the true CDF.

We can use theorem 10 to establish another straightforward result. To do this, we need the following definition.

**Definition 11**

- (a)  $a$  has a positive effect on  $y$  if and only if  $\forall y^0 \in [\underline{y}, \bar{y}]$  and  $a^0 \leq a^1$  with  $a^0, a^1 \in [\underline{a}, \bar{a}] : F_{y^*,a^*}(y^0 | a^0) \geq F_{y^*,a^*}(y^0 | a^1)$ ;
- (b)  $a$  has a negative effect on  $y$  if and only if  $\forall y^0 \in [\underline{y}, \bar{y}]$  and  $a^0 \leq a^1$  with  $a^0, a^1 \in [\underline{a}, \bar{a}] : F_{y^*,a^*}(y^0 | a^0) \leq F_{y^*,a^*}(y^0 | a^1)$ .

We say that  $a$  has a positive effect on  $y$  if the distribution of  $y$ , conditional on  $a$  is non increasing in  $a$ . This means that a higher value for  $a$  gives rise to a distribution of  $y$  that first order stochastically dominates the distributions of  $y$  associated with lower values of  $a$ . Suppose that  $\underline{u} < \bar{u} \leq 0$ , such that  $a^0 - u \geq a^0$ . If  $a$  has a positive (negative) effect on  $y$ , then  $F_{y^*,a^*}(y^0 | a^0 - u) \leq (\geq) F_{y^*,a^*}(y^0 | a^0)$  and  $\Delta^a(y^0 | a^0) \leq (\geq) 0$ . If, on the other hand,  $0 \leq \underline{u} < \bar{u}$ , then if  $a$  has a positive (negative) effect on  $y$ ,  $F_{y^*,a^*}(y^0 | a^0 - u) \geq (\leq) F_{y^*,a^*}(y^0 | a^0)$  and  $\Delta^a(y^0 | a^0) \geq (\leq) 0$ . This already shows in a simple way the importance of the sign of the effect of  $a$  on  $y$  for the effect of measurement error in  $a$  on the conditional cumulative distribution function of  $y$ .

The result of theorem 10 shows that  $\Delta^a(y^0 | a^0)$  is proportional to a weighted average of deviations of  $F_{y^*,a^*}(y^0 | a^0 - u)$  from  $F_{y^*,a^*}(y^0 | a^0)$ , where the weights are given by

$$h_{a^*,u}(a^0, u) \equiv f_{a^*}(a^0 - u) f_u(u) \quad (4)$$

$h_{a^*,y}(a^0, u)$  is the density of  $(a^0, u)$ . We can show the following theorem.

#### Corollary 12

- (a1) If  $a$  has a positive effect on  $y$ ,  $h_{a^*,y}(a^0, u) \in \{f_{a^*,u}(a^0, u) \in \Psi^2 | f_{a^*,u}(a^0, u) \text{ is PRD in } u \text{ around } 0\}$  and  $F_{y^*,a^*}(y | a_0) \in D_a^{AL}$  then  $\Delta^y(y^0 | a^0) \geq 0$ .
- (a2) If  $a$  has a positive effect on  $y$ ,  $h_{a^*,y}(a^0, u) \in \{f_{a^*,u}(a^0, u) \in \Psi^2 | f_{a^*,u}(a^0, u) \text{ is PLD in } u \text{ around } 0\}$  and  $F_{y^*,a^*}(y | a_0) \in D_a^{AR}$  then  $\Delta^y(y^0 | a^0) \leq 0$ .
- (b1) If  $a$  has a negative effect on  $y$ ,  $h_{a^*,y}(a^0, u) \in \{f_{a^*,u}(a^0, u) \in \Psi^2 | f_{a^*,u}(a^0, u) \text{ is PRD in } u \text{ around } 0\}$  and  $F_{y^*,a^*}(y | a^0) \in D_a^{AR}$  then  $\Delta^y(y^0 | a^0) \geq 0$ .
- (b2) If  $a$  has a negative effect on  $y$ ,  $h_{a^*,y}(a^0, u) \in \{f_{a^*,u}(a^0, u) \in \Psi^2 | f_{a^*,u}(a^0, u) \text{ is PLD in } u \text{ around } 0\}$  and  $F_{y^*,a^*}(y | a^0) \in D_a^{AL} : \Delta^y(y^0 | a^0) \leq 0$ .

The sign of the effect of  $a$  on  $y$  determines the sign of  $F_{y^*,a^*}(y^0 | a^0 - u) - F_{y^*,a^*}(y^0 | a^0)$  for  $u > 0$  and  $u < 0$ . Whether  $F_{y^*,a^*}(y | a)$  is in  $D_a^{AL}$  or in  $D_a^{AR}$  determines which of the two ( $u > 0$  or  $u < 0$ ) is largest in absolute value. The assumption on  $h_{a^*,u}(a^0, u)$  ensures that the deviation with the largest absolute value gets the largest weight.

It is straightforward to formulate sufficient conditions on the distributions of  $f_{a^*}(a)$  and  $f_u(u)$  that make  $h_{a^*,u}(a^0, u)$  pointwise left dominated, right dominated or symmetric. That is done in the following corollary.

**Corollary 13**

(a) If  $f_u(u)$  is PLD around  $\hat{u}$  and  $f_{a^*}(a)$  is PRD around  $a^0 - \hat{u}$ , then  $h_{a^*,u}(a^0, u)$  is PLD in  $u$  around  $\hat{u}$ ;

(b) If  $f_u(u)$  is PRD around  $\hat{u}$  and  $f_{a^*}(a)$  is PLD around  $a^0 - \hat{u}$ , then  $h_{a^*,u}(a^0, u)$  is PRD in  $u$  around  $\hat{u}$ ;

(c) If  $f_u(u)$  is symmetric around  $\hat{u}$  and  $f_{a^*}(a)$  is symmetric around  $a^0 - \hat{u}$ , then  $h_{a^*,u}(a^0, u)$  is symmetric in  $u$  around  $\hat{u}$ .

Since symmetric distributions are both pointwise left and right dominant it is clear that the following holds true:

**Corollary 14**

$\forall h_{a^*,y}(a^0, u) \in \{f_{a^*,u}(a^0, u) \in \Psi^2 \mid f_{a^*,u}(a^0, u) \text{ is PS in } u \text{ around } 0\}$

(a1) If  $a$  has a positive effect on  $y$  and  $F_{y^*,a^*}(y \mid a^0) \in D_a^{AL}$  then  $\Delta^y(y^0 \mid a^0) \geq 0$ .

(a2) If  $a$  has a positive effect on  $y$  and  $F_{y^*,a^*}(y \mid a^0) \in D_a^{AR}$  then  $\Delta^y(y^0 \mid a^0) \leq 0$ .

(b1) If  $a$  has a negative effect on  $y$  and  $F_{y^*,a^*}(y \mid a^0) \in D_a^{AR}$  then  $\Delta^y(y^0 \mid a^0) \geq 0$ .

(b2) If  $a$  has a negative effect on  $y$  and  $F_{y^*,a^*}(y \mid a^0) \in D_a^{AL}$  then  $\Delta^y(y^0 \mid a^0) \leq 0$ .

The comparison of corollaries 14 and 12 yields a similar lesson to the comparison between corollaries 6 and 5. Even if the sign of the effect of  $a$  on  $y$  is determined and we know that  $F_{y^*,a^*}(y \mid a)$  is in  $D_a^{AL}$  or is in  $D_a^{AR}$ , we cannot be sure about the sign of  $\Delta^a(y^0 \mid a^0)$ . A result that does not depend on the sign of the effect of  $a$  on  $y$  can be derived if  $F_{y^*,a^*}(y^0 \mid a)$  is APS in  $a$  around  $a_0$ :

**Corollary 15** If  $h_{a^*,u}(a^0, u)$  is symmetric in  $u$  around 0, and  $F_{y^*,a^*}(y^0 \mid a)$  is APS in  $a$  around  $a^0$  over  $[a^0 - \bar{u}, a^0 - \underline{u}]$  then  $\Delta^a(y^0 \mid a^0) = 0$

Again, it may be possible to rule out certain cases described in corollaries 14 and 12 if auxillary data is available to test whether  $f_u(u)$  or  $f_{a^*}(a)$  are pointwise left or right dominant around zero.

To derive a local result we can take a second order approximation to the expression on the right hand side of Theorem 10. Since the first and second moments of  $u$  are  $m_u^1$  and  $m_u^2$ , we can show

**Corollary 16** *Define*

$$A(m^1, m^2) = \frac{1}{f_{a^*}(a^0) - f'_{a^*}(a^0)m_u^1 + \frac{1}{2}f''_{a^*}(a^0)m_u^2}$$

$$B = - \left[ \frac{\partial F_{y^*, a^*}(y^0 | a^0)}{\partial a^*} f_{a^*}(a^0) \right]$$

$$C = + \frac{1}{2} \left[ \frac{\partial^2 F_{y^*, a^*}(y^0 | a^0)}{(\partial a^*)^2} f_{a^*}(a^0) + 2 \frac{\partial F_{y^*, a^*}(y^0 | a^0)}{\partial a^*} f'_{a^*}(a^0) \right]$$

then

$$\Delta^a(y^0 | a^0) \simeq A(m_u^1, m_u^2) [Bm_u^1 + Cm_u^2]$$

Obviously if  $y^*$  and  $a^*$  are independent then  $F_{y^*, a^*}(y^0 | a)$  does not vary with  $a$ . In that case both  $B$  and  $C$  are zero and  $\Delta^a(y^0 | a^0) \approx 0$ . If  $f_{a^*}(a)$  is uniformly distributed, the sign of  $B$  is opposite to the sign of  $\frac{\partial F_{y^*, a^*}(y^0 | a^0)}{\partial a^*}$  and the sign of  $C$  is equal to the sign of  $\frac{\partial^2 F_{y^*, a^*}(y^0 | a^0)}{(\partial a^*)^2}$ . The sign of this expression depends on whether the true conditional cumulative distribution function is a concave or convex function at  $a^* = a^0$ . If a second order approximation to either of the two functions  $f_{a^*}(a)$  or  $F_{y^*, a^*}(y | a)$  does not suffice, then the higher moments of the distribution of  $u$  will play a non negligible role.

## 2.4 Omitted conditioning variable

Suppose now that we want to estimate the cumulative distribution of  $Y$ , conditional upon  $A$  and  $B$  where both  $Y$  and  $A$  are observed but  $B$  is not. Omitting  $B$  in the computation of the conditional cumulative density means that we compute  $F_{y^*, a^*}(y^0 | a^0)$  instead of  $F_{y^*, a^*, b^*}(y^0 | a^0, b)$ ,  $\forall b \in [\underline{b}, \bar{b}]$ . The following theorem gives the relationship between the former and the latter.

**Theorem 17** *Omitted conditioning variable*

$$F_{y^*, a^*}(y^0 | a^0) = \int_{\underline{b}}^{\bar{b}} F_{y^*, a^*, b^*}(y^0 | a^0, \tilde{b}) f_{b^*, a^*}(\tilde{b} | a^0) d\tilde{b}$$

The theorem says that the misspecified conditional cumulative density  $F_{y^*, a^*}(y^0 | a^0)$  is a weighted average of the true conditional cumulative densities  $F_{y^*, a^*, b^*}(y^0 | a^0, \tilde{b})$ . The weights are given by the conditional densities  $f_{b^*, a^*}(\tilde{b} | a^0)$ . To analyse the influence of omitted variables on the

computed CDF, we have to compare  $F_{y^*,a^*}(y^0 | a^0)$  to a benchmark. One possibility would be to fix a value for  $b$ , say  $b^R$  and compare  $F_{y^*,a^*}(y^0 | a^0)$  to  $F_{y^*,a^*,b^*}(y^0 | a^0, b^R)$ . Another natural comparison is to compare  $F_{y^*,a^*}(y^0 | a^0)$  to the mean of the distributions  $F_{y^*,a^*,b^*}(y^0 | a^0, \tilde{b})$ , defined as

$$F^R(y^0 | a^0) = \int_{\underline{b}}^{\bar{b}} F_{y^*,a^*,b^*}(y^0 | a^0, \tilde{b}) f_{b^*}(\tilde{b}) d\tilde{b}$$

The following corollary establishes results for these cases:

**Corollary 18**

$$\begin{aligned} (a) & F_{y^*,a^*}(y_0 | a_0) - F_{y^*,a^*,b^*}(y_0 | a_0, b^R) = \\ & F_{y^*,a^*,b^*}(y_0 | a_0, \bar{b}) - F_{y^*,a^*,b^*}(y_0 | a_0, b^R) - \int_{\underline{b}}^{\bar{b}} \left| \frac{\partial F_{y^*,a^*,b^*}(y_0 | a_0, b)}{\partial b} \right|_{b=\tilde{b}} F_{b^*,a^*}(\tilde{b} | a_0) d\tilde{b} \\ (b) & F_{y^*,a^*}(y^0 | a^0) - F^R(y^0 | a^0) = - \int_{\underline{b}}^{\bar{b}} \left| \frac{\partial F_{y^*,a^*,b^*}(y^0 | a^0, b)}{\partial b} \right|_{b=\tilde{b}} \left[ F_{b^*,a^*}(\tilde{b} | a^0) - F_{b^*}(\tilde{b}) \right] d\tilde{b} \end{aligned}$$

Corollary 18 examines the consequences of omitting a conditioning variable for estimating the conditional distribution functions. If  $\frac{\partial F_{y^*,a^*,b^*}(y^0 | a^0, b)}{\partial b} = 0$ , that is  $F_{y^*,a^*,b^*}(y^0 | a^0, b)$  is independent from  $b$  then  $F_{y^*,a^*}(y^0 | a^0) = F_{y^*,a^*,b^*}(y^0 | a^0, b)$ ,  $\forall b \in [\underline{b}, \bar{b}]$  since  $F_{y^*,a^*,b^*}(y_0 | a_0, \bar{b}) - F_{y^*,a^*,b^*}(y_0 | a_0, b^R)$ . Furthermore  $F_{y^*,a^*}(y^0 | a^0) = F^R(y^0 | a^0)$ .

Recall that  $\frac{\partial F_{y^*,a^*,b^*}(y^0 | a^0, b)}{\partial b} < (>) 0$  means that  $b$  has a positive (negative) effect on  $y$ . Then if  $b$  has a positive effect on  $y$  Theorem 18 (a) can be used to show that the bias takes on its largest positive value when the reference value of  $b$  ( $b^R$ ) is  $\bar{b}$ .

Theorem 18 (b) can be used to establish the sign of  $F_{y^*,a^*}(y^0 | a^0) - F^R(y^0 | a^0)$  once we note that a transformation of  $f_{b^*,a^*}(b, a^0)$  which decreases the covariance between  $b$  and  $a$  decreases the value of  $F_{b^*,a^*}(b | a^0)$  for low values of  $a^0$  and increases this value for high values of  $a^0$  (see, e.g., Atkinson and Bourguignon (1982)). Since  $F(b)$  is a weighted average of the  $F(b|a)$ , we know that there exists for  $F(b|a)$  a value for  $a$ , say  $\hat{a}$  such that  $F(b|\hat{a}) = F(b)$ . For values below  $\hat{a}$  we have that  $F(b|a) \geq F(b|\hat{a}) = F(b)$  and for values above  $\hat{a}$ ,  $F(b|a) \leq F(b|\hat{a}) = F(b)$ . Now, the value of  $\hat{a}$  for which  $F(b|\hat{a}) = F(b)$  will in general depend on the level of  $b$ . Define low values of  $a$  as values that are below the lowest  $\hat{a}(b)$ , and high values of  $a$  as values greater than the highest value for  $\hat{a}(b)$ . Table 1 gives the sign of  $F_{y^*,a^*}(y^0 | a^0) - F^R(y^0 | a^0)$ .

Keeping the underlying true conditional distributions,  $F_{y^*,a^*,b^*}(y^0 | a^0, b)$ ,  $\forall b \in [\underline{b}, \bar{b}]$  fixed, we can conclude the following. Starting from a situation

in which there is a positive covariance between  $a$  and  $b$ , a decrease in the covariance will bring  $F_{y^*a^*}(y^0 | a^0)$  closer to  $F^R(y^0 | a^0)$ .

A further result can be obtained by assuming symmetry of the distribution  $f_{b^*,a^*}(b | a^0)$ .

**Corollary 19** *Let  $\mu \equiv E(y | a^0, E(b | a^0))$ .*

*If  $f_{b^*,a^*}(b | a^0)$  is PS in  $b$  around  $E(b | a^0)$  and*

*$\frac{\partial F_{y^*,a^*,b^*}(\mu | a^0, b)}{\partial b}$  is PS in  $b$  around  $E(b | a^0)$ , then*

*$F_{y^*,a^*}(\mu | a^0) = F_{y^*,a^*,b^*}(\mu | a^0, E(b | a^0))$*

The assumption that  $\frac{\partial F_{y^*,a^*,b^*}(\mu | a^0, b)}{\partial b}$  is PS in  $b$  around  $E(b | a^0)$  means that the effect of a change in  $b$  on  $F_{y^*,a^*,b^*}(\mu | a^0, b)$  is the same for values of  $b$  whose distance to  $E(b | a^0)$  is the same. If the variables are jointly normally distributed, then this assumption is satisfied. This corollary plays an important role in the interpretation of our findings for the intergenerational model.

### 3 Intergenerational Mobility Model: An Application

In this section of the paper we simulate a model of intergenerational mobility in which the key parameters are calibrated so as to mirror the results of recent studies. We then examine the consequences of measurement error for this model. We consider a simple reduced form model of intergenerational mobility (see, e.g., Conlisk (1974) and Becker and Tomes (1979, 1986)):

$$y_t^* = \beta y_{t-1}^* + \gamma b_t + e_t^y \tag{5}$$

$$b_t = \rho b_{t-1} + e_t^b \tag{6}$$

where  $e_t^y$  and  $e_t^b$  are iid with mean zero and variance  $\sigma_{ey}^2$  and  $\sigma_{eb}^2$ , respectively.  $y_t^*$  denotes a child's permanent income,  $y_{t-1}^*$  denotes parent's

permanent income and  $b_t$  denotes endowments. Both  $y_t^*$  and  $y_{t-1}^*$  are expressed as deviations from their mean. The son's endowments are positively related to his father's endowments. We assume that  $b_t$  follows a first order autoregressive process, (6), where  $\rho$  represents the degree of inheritability of endowments.

In the following two subsections we calibrate this simple intergenerational mobility model. We first assume that  $\gamma = 0$  to concentrate on the measurement error issue in the earnings equation (5) and we fix  $\beta = 0.5$ . We assume that  $e_t^y$  is normal with mean 0 and variance  $\sigma_{e_y}^2$  and is uncorrelated with  $y_{t-1}^*$ . We assume that  $y_t^*$  and  $y_{t-1}^*$  come from a stationary distribution with mean zero and variance  $\sigma_y^2$ . From equation (6) with  $\gamma = 0$ ,  $\sigma_y^2 = \frac{1}{1-\beta^2} \sigma_{e_y}^2$ . We choose  $\sigma_y^2 = .3$  ( $\sigma_y = .55$ ), which is close to the variance of earnings in the NLS as reported by Zimmerman (1992).

Assuming particular distributions for the stochastic components is simply a convenience which allows us to obtain analytic solutions for this model. The theorems presented in the earlier section require no such parametric assumptions. In what follows we use these specific distributions to illustrate the theorems outlined in Section 2. Section 4 looks at the implications of these theorems for estimates of mobility.

### 3.1 Measurement Error in the Child's Income

In this section, we use the model outlined above to illustrate the consequences of measurement error in the child's income for the conditional distributions. The true model as outlined above in (5) with  $\gamma = 0$  is given by:

$$y_t^* = \beta y_{t-1}^* + e_t^y \quad (7)$$

Unfortunately, we cannot observe  $y_t^*$ . Instead we observe  $y_t$  which is defined as:

$$y_t = y_t^* + u_t \quad (8)$$

We assume that  $u_t$  is normally distributed with mean 0 and variance  $\sigma_u^2$  and that  $u_t$  is independent from  $y_t^*$  and  $e_t^y$ . Under these assumptions, the following corollary gives the true distribution  $F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  and the computed distribution  $F_{y_t, y_{t-1}^*}(y_t | y_{t-1}^0)$ .



### Corollary 20

- (a)  $F_{y_t^*, y_{t-1}^*}(y_t^0 | y_{t-1}^0) = F_e(y_t^0 - \beta y_{t-1}^0)$   
(b)  $F_{y_t, y_{t-1}^*}(y_t^0 | y_{t-1}^* = y_{t-1}^0) = \Pr\{e_t^y - u_t \leq y_t^0 - \beta y_{t-1}^0\}$

Given that  $f_e(e)$  is normally distributed with mean zero and variance  $\sigma_{ey}^2$ ,  $f_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  will be a normal distribution with mean  $\beta y_{t-1}^0$  and variance  $\frac{1}{1-\beta^2}\sigma_{ey}^2$ . Since  $(e_t^y - u_t)$  is distributed normally with mean 0 and variance  $(\sigma_e^2 + \sigma_u^2)$ ,  $f_{y_t, y_{t-1}^*}(y_t | y_{t-1}^0)$  is a normal distribution with mean  $\beta y_{t-1}^0$  and variance  $\frac{1}{1-\beta^2}(\sigma_{ey}^2 + \sigma_u^2)$ .

Having obtained analytical solutions, we can relate them to the results of section 2.1, corollaries 6 and 7. We assumed that  $u_t$  is normal with mean zero. Therefore  $f_u(u)$  is symmetric around zero. Also, since the conditional distribution is normal,  $f_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  is symmetric around  $E(y_t^* | y_{t-1}^0)$ . From corollary 7 we know that the distributions from the true and misspecified model must coincide at this point i.e.  $\Delta^y(E\{y_t^* | y_{t-1}^1\} | y_{t-1}^1) = 0$ . However, for any  $y$  value other than  $E\{y_t^* | y_{t-1}^1\}$ ,  $\Delta^y(y_t | y_{t-1}^1) \neq 0$ . To the left of  $E\{y_t^* | y_{t-1}^1\}$  the true CDF is (locally) convex in  $y$ . If the weights attached to this convex part are large enough, then we will have  $\Delta^y(y_t | y_{t-1}^1) > 0$  for  $y_t < E\{y_t^* | y_{t-1}^1\}$ . To the right of  $E\{y_t^* | y_{t-1}^1\}$  the true CDF is (locally) concave, such that we get the opposite result if the weight attached to the concave part is large enough.

This is illustrated in Figure ??, which plots the misspecified and true cdfs, conditional on the median level of father's income. Here the true and misspecified cdf's cross at  $E\{y_t^* | y_{t-1}^1\}$  which is zero. We obtain similar results when we condition on other levels of father's income. One immediate consequence of this is that while the conditional mean of both distributions is the same, the conditional variance of the misspecified distribution is larger than the true distribution. This is to be expected given the extra noise induced by the measurement error. This will have a consequence for estimating mobility which we discuss later in the paper.

## 3.2 Measurement Error in Father's Income

In this section we examine the consequences of measurement error in father's incomes for the calculated distributions. We modify the model of the previous section to examine a situation where the true value of son's income  $y_t^*$  is observed but father's income is measured with error. We specify the observed value of father's income as

$$y_{t-1} = y_{t-1}^* + v_t \quad (9)$$

We assume that  $v_t$  is normally distributed with mean zero and variance  $\sigma_v^2$ . The assumptions concerning the error term are analogous to those made in the previous section, namely that it is normally distributed and independent of  $y_{t-1}^*$ . The true underlying distributions are identical to those presented in corollary 20 (a). In this case it is difficult to obtain a recognizable closed form expression for the misspecified distribution,  $F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0) = \Pr\{y_t^* \leq y_t | y_{t-1} = y_{t-1}^0\}$ . We use theorem 10 to evaluate the density at any chosen value for  $y_t^*$  and  $y_{t-1}^*$ . Note that  $F_{y_t^*, y_{t-1}^*}(0 | y_{t-1}^*)$  is symmetrical in  $y_{t-1}^*$  around 0, that  $f_{y_{t-1}^*}(y_{t-1})$  is symmetrical around 0 and that  $f_u(u)$  is symmetrical around 0. This implies, by corollary 2 and 15, that  $\Delta^a(0 | 0) = 0$ .

The results are given in Figures 3-5, for estimated densities conditional on the 25th percentile, the median and the 75th percentile of the respective distributions. Figure 4 confirms that  $\Delta^a(0 | 0) = 0$ . The figures show, that for our specification, the effect of measurement error is smaller when we condition on the median value of father's income than when we condition on the values of father's income that are further away from the median. Furthermore the direction of the bias switches at the median. In particular when we condition on the 25th percentile of father's income we substantially underestimate the true distribution (Figure 3), whereas conditioning on values above the median of father's income leads us to overestimate the true distribution at all values of  $y_t^*$  (Figure 5).<sup>6</sup>

The misspecified distributions would lead us to believe that children of poorer fathers appear wealthier than they actually are, while children of richer parents appear poorer than they actually are. This result incorporates the traditional bias in the linear regression model, where it is well known that measurement error in a control variable tends to bias the estimated coefficient towards zero. However our results are more general in that they imply that not only is the slope of the mean function biased towards zero but so too is the slope of any quantile function estimated with the contaminated data. These findings have important implications for patterns of mobility which are discussed in section 4.

---

<sup>6</sup>These results were used by Friedman (1957) to motivate his theory of permanent income. For a discussion of the intuition behind these findings see Friedman (1992).

### 3.3 Omitted Ability

To study the influence of omitted ability on the position of  $F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  relative to the distributions in  $F_{y_t^*, y_{t-1}^*, b_t}(y_t | y_{t-1}^0, b_t), \forall b_t \in [\underline{b}, \bar{b}]$ , we have to consider the complete intergenerational model, given by equations (5) and (6). Appendix 2 contains the solution of this intergenerational model. To obtain the graphs in this section we assumed that  $\beta = \gamma = \rho = 0.5$  and  $\sigma_y^2 = \sigma_b^2$ . Figure 6 compares the position of  $F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  to the position of three curves  $F_{y_t^*, y_{t-1}^*, b_t}(y_t | y_{t-1}^0, b_t)$ , corresponding to the 2.5-th percentile,  $E(b_t | y_{t-1}^0)$  and the 97.5-th percentile of the distribution of  $b_t$ , with  $y_{t-1}^0$  equal to its mean value. Figures 7 and 8 do the same for  $y_{t-1}^0$  equal to the 2.5-th and 97.5-th percentile of the distribution of  $y_{t-1}$ .

Equation (5) specifies a linear relationship between  $y_t^*, y_{t-1}^*$  and  $b_t$ . Consequently, the cumulative conditional distributions  $F_{y_t^*, y_{t-1}^*, b_t}(y_t | y_{t-1}^0, b_t)$  are simple horizontal shifts of the same cumulative distribution function that is determined by the distribution of  $e_t^y$ . With normality of the latter distribution, all conditional distributions will be normal. Consequently,  $f_{b_t, y_{t-1}^*}(b_t | y_{t-1}^0)$  is symmetric in  $b_t$  around  $E(b_t | y_{t-1}^0)$  and  $\frac{\partial F_{y_t^*, y_{t-1}^*, b_t}(\mu | y_{t-1}^0, b_t)}{\partial b_t}$  will be symmetric in  $b_t$  around  $E(b_t | y_{t-1}^0)$ . Corollary 19 therefore states that

$F_{y_t^*, y_{t-1}^*}(\mu | y_{t-1}^0) = F_{y_t^*, y_{t-1}^*, b_t}(\mu | y_{t-1}^0, E(b_t | y_{t-1}^0))$ . This explains why in all the graphs these two curves cross at  $y_t = \mu = E(y_t | y_{t-1}^0, E(b_t | y_{t-1}^0))$  and both functions are equal to 1/2 at that point.

Next consider  $y_t^0 < E(y_t | y_{t-1}^0, E(b_t | y_{t-1}^0))$ . Due to the normality assumptions,  $\frac{\partial F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, b_t)}{\partial b_t}$  will be symmetric around a value for  $b_t$ , say  $\hat{b}_t$ . Since  $\frac{\partial F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, b_t)}{\partial b_t} < 0$  we have that  $\hat{b}_t < E(b_t | y_{t-1}^0)$ . Normality also implies that  $F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, b_t)$  will be a convex function of  $b_t$  in the neighborhood of  $b_t = E(b_t | y_{t-1}^0)$ .

In (16) this convexity implies that

$$\begin{aligned} F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, E(b_t | y_{t-1}^0) + c) + F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, E(b_t | y_{t-1}^0) - c) \\ > 2F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, E(b_t | y_{t-1}^0)) \end{aligned}$$

which in turn implies that

$$F_{y_t^*, y_{t-1}^*}(y_t^0 | y_{t-1}^0) > F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, E(b_t | y_{t-1}^0))$$

This explains why  $F_{y_t^*, y_{t-1}^*}(y_t^0 | y_{t-1}^0)$  lies above  $F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, E(b_t | y_{t-1}^0))$

in figures 6-8 for  $y_t^0 < E(y_t | y_{t-1}^0, E(b_t | y_{t-1}^0))$ . The opposite conclusion holds for  $y_t^0 > E(y_t | y_{t-1}^0, E(b_t | y_{t-1}^0))$  because  $F_{y_t^*, y_{t-1}^*, b_t}(y_t^0 | y_{t-1}^0, b_t)$  will be a concave function of  $b_t$  in the neighborhood of  $b_t = E(b_t | y_{t-1}^0)$ .

## 4 Evaluating the consequences of mis specification for the Measurement of Mobility

In this section we illustrate the consequences of these specification errors for measures of mobility. Since mobility can have many meanings (see, e.g., Van de gaer, Schokkaert and Martinez (2001)), we must begin by defining exactly what we mean by it in this context. An important tradition in economics and sociology measures mobility on the basis of transition matrices (see, e.g., Boudon (1973), Shorrocks (1978), Dardanoni (1993) and Conlisk (1990)). It is natural to think of mobility decreasing as more children stay in the same income quantile as their father.

The mobility measure we use is in the same spirit as the immobility ratio that Atkinson et al (1992) discuss in the context of transition matrices. For example, if we consider children of parents who are at the 20th percentile of the parental distribution, we class this child as not moving if he is within a suitably defined interval of the 20th percentile of the children's income distribution. Formally, the immobility range is determined by  $\delta^l(\alpha, y_{t-1})$  and  $\delta^u(\alpha, y_{t-1})$  which are defined for a chosen probability  $\alpha$ , so that:

$$\delta^u(\alpha, y_{t-1}) \longleftrightarrow F_{y_t^*}(y_a + \delta^u(\alpha, y_{t-1})) - F_{y_t^*}(y_a) = \alpha \quad (10)$$

and

$$\delta^l(\alpha, y_{t-1}) \longleftrightarrow F_{y_t^*}(y_a) - F_{y_t^*}(y_a - \delta^l(\alpha, y_{t-1})) = \alpha \quad (11)$$

where  $y_a$  is defined such that  $F_{y_t^*}(y_a) = F_{y_{t-1}^*}(y_{t-1})$ .

We can calculate a true measure of mobility  $M(\alpha, y_{t-1})$  by calculating the probability that the income of the son will fall outside these bands in the conditional income distribution. That is:

$$M(\alpha, y_{t-1}) = 1 - \left( F_{y_t^*, y_{t-1}^*}(y_a + \delta^u(\alpha, y_{t-1}) | y_{t-1}) - F_{y_t^*, y_{t-1}^*}(y_a - \delta^l(\alpha, y_{t-1}) | y_{t-1}) \right) \quad (12)$$

We can compare this to  $M^e(\alpha, a)$  which represents mobility calculated from the density measured with error. This is calculated as:

$$M^e(\alpha, y_{t-1}) = 1 - (F^e(y_a + \delta^u(\alpha, y_{t-1}) | y_{t-1}) - F^e(y_a - \delta^l(\alpha, y_{t-1}) | y_{t-1})) \quad (13)$$

By comparing the function defined by (12) with the function defined by (13) as a function of  $y_{t-1}$ , we are able to see where in the distribution we over or underestimate mobility. In practice, we do this for nine points of the distribution, the 1st to the 9th income decile. Drawing this graph for several values of  $\alpha$  allows us to see whether the results are sensitive to the width of the bands chosen.

Table 2 presents the true mobility index,  $M(\alpha, y_{t-1})$ , and the mobility index when son's income and father's income is measured with error,  $M^e(\alpha, y_{t-1})$ . We report the indices for the nine specified values of the parents' income and set  $\alpha$  equal to .1. Thus we define an individual as having moved if their ranking in the children's income distribution is at least ten percentage points higher or lower than their father's ranking. For example the value of .71 appearing in the second row of Table 2, tells us that conditional on their father having an income level corresponding to the 20th percentile of the parental distribution, 71% of children will be ranked either below the 10th percentile or above the 30th percentile in the children's income distribution. For both the true and misspecified distributions we see that given our distributional assumptions mobility is higher in the middle of the distribution than in the tails. However, what is of importance for us in this paper is the comparison of the mobility measures under the true and misspecified models. Our results show that when the child's or father's income is measured with error, the proportion of individuals classified as moving rises. Given our specifications it is also the case that the extent to which mobility is overestimated seems to be higher in the tails than in the middle of the distribution.

Table 3 analyses the consequences of omitted ability for the computation of our mobility index. In doing so we have to compare the two dimensional process  $F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  to the three dimensional process  $F_{y_t^*, y_{t-1}^*, b_t}(y_t | y_{t-1}^0, b_t)$ . To do this we choose three alternative reference values  $b^R$  corresponding to the 25th, 50th and 75th percentile value for  $b_t$ . The results in Table 3 compare the misspecified measure of mobility  $M^e(.1, y_{t-1})$ , based on

$F_{y_t^*, y_{t-1}^*}(y_t | y_{t-1}^0)$  (column 1), to the true measure of mobility  $M(.1, y_{t-1}, b^R)$  based on  $F_{y_t^*, y_{t-1}^*, b_t^*}(y_t | y_{t-1}^0, b^R)$  for each of the three values of  $b^R$  (columns 2-4). Depending on the value of  $b^R$  we see that mobility can either rise monotonically with income (for low ability families,  $b_t = -.37$ ), fall monotonically (for high ability families,  $b_t = .37$ ), or remain approximately constant as income changes (average ability families,  $b_t = 0$ ). However none of these trends are replicated in the misspecified model, where mobility first rises with fathers income and then falls.

The reasons for this pattern can be traced back to figures 6-8. Our mobility measure is based on an interval around an income level in the son's distribution which is determined by their father's income. Due to the assumed stationarity of the distributions in our model this level of income is equal to the father's income. For  $y_{t-1} = 0$  and  $b^R = 0$ , illustrated in figure 5, this interval is defined around  $y_t = 0$ .  $M(\cdot, 0, 0)$  computes the probability mass outside of this interval on the basis of the central dotted CDF.  $M^e(\cdot, 0)$  calculates the probability mass outside of this interval on the basis of the CDF drawn with a solid line. As can be seen from the graph,  $M^e(\cdot, 0) > M(\cdot, 0, 0)$  since the slope of the full CDF is flatter than the slope of the dotted CDF. This effect reflects the additional variance in  $y_t$  caused by the omitted variable in equation (5). To see what happens for other values of fathers income, keeping  $b^R = 0$ , consider figure 7. Our mobility measure constructs an interval around  $y_t = -1.074$ .  $M^e(\cdot, -1.074)$  is computed on the basis  $F_{y_t^*, y_{t-1}^*}(y_t | -1.074)$ , the CDF drawn in full.  $M(\cdot, -1.074, 0)$  is calculated on the basis of  $F_{y_t^*, y_{t-1}^*, b}(y_t | -1.074, 0)$ , a CDF that is situated somewhat to the right of the central dotted CDF. Misspecification has two effects. On the one hand, due to the additional variation in the stochastic component of equation (5) when a variable is omitted, the misspecified CDF is flatter than the true CDF's. This increases mobility. On the other hand, the positive covariance ensures that  $E(b | y_{t-1}^0) < 0$  whenever  $y_{t-1}^0 < 0$  which pushes the misspecified CDF to the left. This has a downward effect on mobility since, for  $y_{t-1}^0 < 0$ ,  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{t-1}^0, 0)$  is flatter than  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{t-1}^0, E(b | y_{t-1}^0))$ . The overall effect becomes undetermined. Figure 8 shows that a similar ambiguity arises at the other side of the mean of  $y_{t-1}$ . The further we move towards the tails of the distribution, the more important the second effect.

It is clear that omitted ability can significantly change underlying mobility patterns. Several studies of intergenerational mobility have emphasized

the fact that mobility tends to be lower from the tails of the income distribution than from the centre. However the results presented in Table 3 highlight the problems that misspecification can cause for this type of analysis. The mobility indices based on the misspecified model,  $M^e(.1, y_{t-1})$ , give the impression of a distribution in which mobility is lower in the tails than in the middle and of a society in which children from poor families are remaining poor, children from rich families are staying rich, while there is more mobility among middle class families. However this is not a complete characterization of this society. When we look at the fully specified intergenerational model, including endowments, we see that the extent to which mobility  $M(.1, y_{t-1})$  varies with father's income depends on the level of endowments. For instance among parents of children with average endowments, individuals whose fathers were poor are just as likely to be classified as movers as individuals with middle or upper class fathers.

## 5 Conclusion

Over the last 10 to 15 years there has been a substantial increase in the number of studies looking at the entire joint distribution of the variables of interest rather than focusing on the relationship at the mean. This is often done by estimating conditional distribution functions. In this paper we examine the properties of cumulative conditional distribution functions in the presence of several types of misspecification.

Section 2 establishes the main findings of the paper. The results are in the spirit of dominance results: they give an indication of the effect of misspecification for classes of true (unknown) distribution functions. Consequently, knowledge of a fairly general feature of the distribution function will help to determine the effect of measurement error.

The effect of omitted conditioning variables depends on the sign of the effect of the omitted variable on the conditioned variable and the covariance between the omitted and the present conditioning variable. This is quite similar to the effect of omitted variables in the linear regression case. In all cases of specification error we considered, knowledge of some general features of the causes of specification error (the distribution of measurement error or the sign of the effect of the unobserved variable on the conditioned variable) help to determine the effect of misspecification.

Sections 3 and 4 use a simulated model of intergenerational mobility to

illustrate the findings. Within the context of the intergenerational mobility model we find that measurement error in the child's income causes researchers to overestimate mobility when the mobility indices are based on the entire distribution of income. This is in contrast to OLS estimates of correlation coefficients, that have been the focus of much of the early intergenerational mobility literature, and which remain unbiased even in the presence of measurement error in the child's income. Furthermore we show that omitted variables can seriously distort mobility patterns and that the magnitude of the bias can vary depending on the level of parental income. In particular we show that it is possible for data to give the impression of a society in which most of the mobility is among middle-class children, when this is not the case in the fully specified model. Thus not only is the degree of mobility affected but so is the pattern of mobility throughout the distribution. This finding may have important consequences for how we use transition matrices, both in order to characterize our society and as a means of distinguishing between competing economic models of mobility.



## References

- Abul Naga, R. (2001), Galtonian Regression of Intergenerational Income Linkages: Biased Procedures, a New Estimator and Mean Square Error Comparisons, Sticerd/LSE, Distributional Analysis Research Program, DARP No.53.
- Angrist, J. and A. Krueger (1999), Empirical Methods in Labor Economics, in O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics Vol.3A, Elsevier.
- Atkinson, A.B. and F. Bourguignon (1982), The comparison of Multi-Dimensioned Distributions of Economic Status, Review of Economic Studies, Vol.49, 183-201.
- Atkinson, A.B, F. Bourguignon and C. Morrison (1992), Empirical Studies of Earnings Mobility, Fundamentals of pure and applied Economics 52, Harwood Academic publishers.
- Atkinson, A.B., A.K. Maynard and C.G. Trinder (1983), Parents and Children, London: Heinemann.
- Becker, G. and N. Tomes (1979), An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility, Journal of Political Economy, Vol.87, 1153-1189.
- Becker, G. and N. Tomes (1986), Human Capital and the Rise and Fall of Families, Journal of Labor Economics, Vol.4, S1-S39.
- Boudon, R. (1973), Mathematical Structures of Social Mobility, Amsterdam: Elsevier.
- Burkhauser, R , A. Crews Cutts, M. Daly and S. Jenkins (1999), Testing the Significance of Income Distribution Changes Over the 1980's Business Cycle: A Cross-National Comparison, Journal of Applied Econometrics, Vol. 14, no. 3 (May-June), 253-272.
- Chesher, A. (1991), The Effect of Measurement Error, Biometrika, Vol.78, 451-462.
- Chesher, A. and C. Schluter (2001), Welfare Measurement and Measurement Error, University of Bristol discussion paper.
- Conlisk, J. (1974), Can Equalization of Opportunity Reduce Social Mobility?, American Economic Review, Vol.64, 80-90.
- Conlisk, J. (1990), Monotone Mobility Matrices, Journal of Mathematical Sociology, Vol.15, 173-191.
- Corak, M. and A. Heisz (1999), The Intergenerational Earnings and Income Mobility of Canadian Men, Journal of Human Resources, Vol.34, 504-

533.

Dardanoni, V. (1993), Measuring Social Mobility, *Journal of Economic Theory*, Vol.46, 372-394.

DiNardo, J., N.M. Fortin and T. Lemieux (1996), Labor Market Institutions and the Distribution of Wages 1973-1992: A Semiparametric Approach, *Econometrica*, Vol.64, 1001-1044.

Friedman, M (1957), *A Theory of the Consumption Function*, Princeton N.J.: Princeton University Press.

Friedman, M (1992), Do Old Fallacies Ever Die?, *Journal of Economic Literature*, Vol.30, 2129-2132.

Fuller, W (1987), *Measurement Error Models*, New York: Wiley and Sons.

Han, S. and Mulligan, C.B. (2000), Human Capital, Heterogeneity, and Estimated Degrees of Intergenerational Mobility, NBER Working paper No. 7678.

Griliches, Z and V. Ringstad (1970), Error-In-The-Variables Bias in Non-linear Contexts, *Econometrica*, Vol.38, 368-370

Hausman, J., W.K Newey and J. Powell (1995), Nonlinear Errors in Variables Estimation of Some Engel Curves, *Journal of Econometrics*, Vol.65, 205-233.

Magnac, T. and M. Visser (1999), Transition Models with Measurement Errors, *Review of Economics and Statistics*, Vol.81, 466-474.

O'Neill, D, O. Sweetman and D. Van de gaer (2000), Equality of Opportunity and Kernel Density Estimation: an Application to Intergenerational Mobility, in T.B. Fomby and R.C. Hill (eds.), *Advances in Econometrics*, JAI Press, Vol.14, 259-274.

Quah, D. (1997) Empirics for Growth and Distribution: Polarization and Convergence Clubs, *Journal of Economic Growth*, Vol.2, 27-59.

Shorrocks, A. (1978), The Measurement of Mobility, *Econometrica*, Vol.46, 1013-1024.

Solon, G. (1992), Intergenerational Income Mobility in the United States, *American Economic Review*, Vol.82, 393-408.

Solon, G. (1999), Intergenerational Mobility in the Labor Market, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* Vol.3A, Elsevier, 1761-1800.

Trede, M. (1998a), Making Mobility Visible: a Graphical Device, *Economics Letters*, Vol.59, 77-82.

Trede, M. (1998b), The Age Profile of Mobility Measures: an Application

to Earnings in West Germany, *Journal of Applied Econometrics*, Vol.13, 397-409.

Van de gaer, D., E. Schokkaert and M. Martinez (2001), Three Meanings of Intergenerational Mobility, *Economica*, Vol.68, 519-537.

Zimmerman, D.J. (1992), Regression Towards Mediocrity in Economic Stature, *American Economic Review*, Vol.82, 409-429.

## Appendix 1: Proofs of the results in section 2

### Proof of theorem 3

Since  $y = y^* + u$ , we have

$$f_{y,a^*}(y^0, a^0) = \int_{\underline{u}}^{\bar{u}} f_{y^*,a^*,u}(y^0 - u, a^0, u) du$$

Since  $u$  is distributed independently from  $(y^*, a^*)$

$$f_{y,a^*}(y^0, a^0) = \int_{\underline{u}}^{\bar{u}} f_{y^*,a^*}(y^0 - u, a^0) f_u(u) du$$

such that

$$F_{y,a^*}(y^0 | a^0) = \int_{\underline{u}}^{\bar{u}} F_{y^*,a^*}(y^0 - u | a^0) f_u(u) du$$

At the same time,

$$F_{y^*,a^*}(y^0 | a^0) = \int_{\underline{u}}^{\bar{u}} F_{y^*,a^*}(y^0 | a^0) f_u(u) du$$

Differencing the last two expressions results in equation (2) of the theorem.

### Proof of corollary 5

(2) can be written as

$$\begin{aligned} \Delta^y(y^0 | a^0) &= \int_{\underline{u}}^0 \underbrace{(F_{y^*,a^*}(y^0 - u | a^0) - F_{y^*,a^*}(y^0 | a^0))}_{A} f_u(u) du \\ &\quad + \int_0^{\bar{u}} \underbrace{(F_{y^*,a^*}(y^0 - u | a^0) - F_{y^*,a^*}(y^0 | a^0))}_{B} f_u(u) du \end{aligned} \quad (14)$$

The first term,  $A$ , will be positive, the second term,  $B$ , will be negative. To determine the sign of (14) note that, if  $F_{y^*,a^*}(y | a_0) \in D_1^{AL}$ , then  $A \leq |B|$ , such that, if in addition  $f_u(u)$  is pointwise right dominant around 0, then  $\Delta^y(y_0 | a_0) \leq 0$ . If, on the other hand,  $F_{y^*,a^*}(y | a_0) \in D_1^{AR}$ , then  $A \geq |B|$  and if, in addition,  $f_u(u)$  is pointwise left dominant around 0, then  $\Delta^y(y_0 | a_0) \geq 0$ . In the above proof the absolute pointwise left (right) dominance of  $F_{y^*,a^*}(y | a^0)$  determines the relative magnitudes of the deviations of the CDF's of  $y$  for values of  $u > 0$  and  $u < 0$ . The assumptions on  $f_u(u)$  in corollary 5 ensure that the deviation that has the largest absolute value dominates.

### Proof of corollary 7

If  $F_{y^*,a^*}(y | a^0)$  is APS in  $y$  around  $E(y^* | a^0)$ , then (14) becomes

$$\begin{aligned} \Delta^y (E(y^* | a^0) | a^0) &= \int_0^{\bar{u}} \underbrace{\left( F_{y^*, a^*} (E(y^* | a^0) + u | a^0) - F_{y^*, a^*} (E(y^* | a^0) | a^0) \right)}_A \\ &\quad + \underbrace{\left( F_{y^*, a^*} (E(y^* | a^0) - u | a^0) - F_{y^*, a^*} (E(y^* | a^0) | a^0) \right)}_B f_u(u) du \end{aligned} \quad (15)$$

In (15),  $A = |B|$ , such that, if  $f_u(u)$  is symmetric around 0, then  $\Delta^y (y^0 | a^0) = 0$ .

### Proof of corollary 9

Follows immediately from using a  $n$ -th order Taylor expansion of  $F_{y^*, a^*} (y^0 - u | a^0)$  around  $y^0$ .

### Proof of theorem 10

Since  $a = a^* + u$ , we have that

$$f_{y^*, a} (y^0, a^0) = \int_{\underline{u}}^{\bar{u}} f_{y^*, a^*, u} (y^0, a^0 - u, u) du$$

We assume that  $u$  is not correlated with  $(y^*, a^*)$  and, therefore, we can write for the conditional cumulative density

$$F_{y^*, a} (y^0 | a^0) = \int_{\underline{u}}^{\bar{u}} F_{y^*, a^*} (y^0 | a^0 - u) \frac{f_{a^*}(a^0 - u)}{f_a(a^0)} f_u(u) du$$

and note, that because of the independence between  $u$  and  $a$

$$f_a(a^0) = \int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du$$

We also have that

$$F_{y^*, a^*} (y^0 | a^0) = \int_{\underline{u}}^{\bar{u}} F_{y^*, a^*} (y^0 | a^0) f_u(u) du$$

Collecting the above expressions and differencing results in the theorem.

### Proof of corollary 12

(3) can be written as

$$\begin{aligned} &\frac{1}{\int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du} \\ &\quad \left\{ \underbrace{\int_{\underline{u}}^{\bar{u}} [F_{y^*, a^*} (y^0 | a^0 - u) - F_{y^*, a^*} (y^0 | a^0)] h_{a^*, u} (a^0, u) du}_{F_1} \right. \\ &\quad \left. + \int_0^{\bar{u}} \underbrace{[F_{y^*, a^*} (y^0 | a^0 - u) - F_{y^*, a^*} (y^0 | a^0)] h_{a^*, u} (a^0, u) du}_{F_2} \right\} \end{aligned}$$

If  $a$  has a positive effect on  $y$ , then  $F_2 > 0$  and  $F_1 < 0$ . If, in addition  $F_{y^*, a^*} (y^0 | a) \in D_2^{AL}$ , then  $|F_2| > |F_1|$ . If then  $h_{a^*, u} (a^0, u)$  is pointwise right dominant in  $u$  around 0, then the expression above will be positive. If  $a$

has a positive effect on  $y$ ,  $F_{y^*,a^*}(y^0 | a) \in D_2^{AR}$  and, in addition  $h_{a^*,u}(a^0, u)$  is pointwise left dominant in  $u$  around 0, then the expression above will be negative.

The results when  $a$  has a negative effect on  $y$  can be derived analogously.

### Proof of corollary 13

(a) Applying definition 1 (a) and the definition (4)  $h_{a^*,u}(a^0, u)$  will be pointwise left dominant in  $u$  around  $\hat{u}$  if and only if

$$f_{a^*}(a^0 - \hat{u} + r) f_u(\hat{u} - r) \geq f_{a^*}(a^0 - \hat{u} - r) f_u(\hat{u} + r).$$

A sufficient condition for this to hold true is that  $f_{a^*}(a^0 - \hat{u} + r) \geq f_{a^*}(a^0 - \hat{u} - r)$  and  $f_u(\hat{u} - r) \geq f_u(\hat{u} + r)$ . By definition 1 (b) the first inequality means that  $f_{a^*}(a)$  is pointwise right dominant around  $a^0 - \hat{u}$ . By 1 (a) the second inequality means that  $f_u(u)$  is pointwise left dominant around  $\hat{u}$ .

(b) can be shown analogously to the proof of (a).

(c) From 1 (c) and the definition (4)  $h_{a^*,u}(a^0, u)$  will be symmetric in  $u$  around  $\hat{u}$  if and only if

$$f_{a^*}(a^0 - \hat{u} + r) f_u(\hat{u} - r) = f_{a^*}(a^0 - \hat{u} - r) f_u(\hat{u} + r)$$

A sufficient condition for this to hold true is that  $f_{a^*}(a^0 - \hat{u} + r) = f_{a^*}(a^0 - \hat{u} - r)$  and  $f_u(\hat{u} - r) = f_u(\hat{u} + r)$ . Remembering definition 1 (c), the proof is completed.

### Proof of corollary 15

If  $F_{y^*,a^*}(y^0 | a)$  is APS in  $a$  around  $a^0$ , then  $F_1 = -F_2$ , such that, if  $h_{a^*,u}(a^0, u)$  is symmetric in  $u$  around 0, then  $\Delta^a(y^0 | a^0) = 0$ .

### Proof of corollary 16

We have that

$$\begin{aligned} & \int_{\underline{u}}^{\bar{u}} F_{y^*,a^*}(y^0 | a^0 - u) \frac{f_{a^*}(a^0 - u)}{\int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du} f_u(u) du \\ &= \frac{1}{\int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du} \int_{\underline{u}}^{\bar{u}} F_{y^*,a^*}(y^0 | a^0 - u) f_{a^*}(a^0 - u) f_u(u) du \end{aligned}$$

A second order approximation of  $f_{a^*}(a^0 - u)$  around  $a^0$  results in

$$\frac{1}{\int_{\underline{u}}^{\bar{u}} f_{a^*}(a^0 - u) f_u(u) du} \simeq \frac{1}{f_{a^*}(a^0) - f'_{a^*}(a^0) m_u^1 + \frac{1}{2} f''_{a^*}(a^0) m_u^2}$$

Similarly,

$$\begin{aligned} & \int_{\underline{u}}^{\bar{u}} F_{y^*,a^*}(y^0 | a^0 - u) f_{a^*}(a^0 - u) f_u(u) du \\ & \simeq F_{y^*,a^*}(y^0 | a^0) f_{a^*}(a^0) \\ & - \left[ \frac{\partial F_{y^*,a^*}(y^0 | a^0)}{\partial a^*} f_{a^*}(a^0) + F_{y^*,a^*}(y^0 | a^0) f'_{a^*}(a^0) \right] m_u^1 \end{aligned}$$

$$+ \frac{1}{2} \left[ \frac{\partial^2 F_{y^*,a^*}(y^0|a^0)}{(\partial a^*)^2} f_{a^*}(a^0) + 2 \frac{\partial F_{y^*,a^*}(y^0|a^0)}{\partial a^*} f'_{a^*}(a^0) + F_{y^*,a^*}(y^0|a^0) f''_{a^*}(a^0) \right] m_u^2$$

Collecting terms results in the corollary.

**Proof of theorem 17**

$$\begin{aligned} F_{y^*,a^*}(y^0|a^0) &= \int_{\underline{y}}^{y^0} f_{y^*,a^*}(\tilde{y}|a^0) d\tilde{y} \\ &= \int_{\underline{y}}^{y^0} \frac{f_{y^*,a^*}(\tilde{y},a^0)}{f_{a^*}(a^0)} d\tilde{y} \\ &= \int_{\underline{y}}^{y^0} \frac{\int_{\underline{b}}^{\bar{b}} f_{y^*,a^*,b^*}(\tilde{y},a^0,\tilde{b}) d\tilde{b}}{f_{a^*}(a^0)} d\tilde{y} \end{aligned}$$

Using the fact that:

$$\begin{aligned} f_{y^*,a^*,b^*}(y,a,b) &= f_{y^*,a^*,b^*}(y|a,b) f_{a^*,b^*}(a,b). \\ F_{y^*,a^*}(y^0|a^0) &= \int_{\underline{y}}^{y^0} \frac{\int_{\underline{b}}^{\bar{b}} f_{y^*,a^*,b^*}(y|a,b) f_{a^*,b^*}(a,b)}{f_{a^*}(a^0)} d\tilde{y} \\ &= \int_{\underline{b}}^{\bar{b}} F_{y^*,a^*,b^*}(y^0|a^0,\tilde{b}) f_{b^*,a^*}(\tilde{b}|a^0) d\tilde{b}. \end{aligned}$$

**Proof of corollary 18 (a)**

If we integrate the right hand side of 17 by parts we get:

$$\begin{aligned} F_{y^*,a^*}(y_0|a_0) &= \left[ F_{y^*,a^*,b^*}(y_0|a_0,b) F_{b^*,a^*}(b|a_0) \right]_{\underline{b}}^{\bar{b}} \\ &\quad - \int_{\underline{b}}^{\bar{b}} \left[ \frac{\partial F_{y^*,a^*,b^*}(y_0|a_0,b)}{\partial b} \right]_{b=\tilde{b}} F_{b^*,a^*}(\tilde{b}|a_0) d\tilde{b} \end{aligned}$$

Evaluating the first expression behind the equality sign we get

$$\begin{aligned} F_{y^*,a^*}(y_0|a_0) &= F_{y^*,a^*,b^*}(y_0|a_0,\bar{b}) - \\ &\quad \int_{\underline{b}}^{\bar{b}} \left[ \frac{\partial F_{y^*,a^*,b^*}(y_0|a_0,b)}{\partial b} \right]_{b=\tilde{b}} F_{b^*,a^*}(\tilde{b}|a_0) d\tilde{b}. \end{aligned}$$

Subtracting  $F_{y^*,a^*,b^*}(y_0|a_0,b^R)$  establishes the result.

**Proof of corollary 18 (b)**

Integrating by parts we get:

$$\begin{aligned} F_{y^*,a^*}(y^0|a^0) - F^r(y^0|a^0) &= \left[ F_{y^*,a^*,b^*}(y^0|a^0,b) [F_{b^*,a^*}(b|a^0) - F_{b^*}(b)] \right]_{\underline{b}}^{\bar{b}} \\ &\quad - \int_{\underline{b}}^{\bar{b}} \left[ \frac{\partial F_{y^*,a^*,b^*}(y^0|a^0,b)}{\partial b} \right]_{b=\tilde{b}} \left[ F_{b^*,a^*}(\tilde{b}|a^0) - F_{b^*}(\tilde{b}) \right] d\tilde{b} \end{aligned}$$

Evaluating we get

$$\begin{aligned} F_{y^*,a^*}(y^0|a^0) - F^r(y^0|a^0) &= \\ &\quad - \int_{\underline{b}}^{\bar{b}} \left[ \frac{\partial F_{y^*,a^*,b^*}(y^0|a^0,b)}{\partial b} \right]_{b=\tilde{b}} \left[ F_{b^*,a^*}(\tilde{b}|a^0) - F_{b^*}(\tilde{b}) \right] d\tilde{b}. \end{aligned}$$

**Proof of corollary 19**

From theorem 17, we have that

$$F_{y^*, a^*}(y^0 | a^0) = \int_{\underline{b}}^{\bar{b}} F_{y^*, a^*, b^*}(y^0 | a^0, b) f_{b^*, a^*}(b | a^0) db$$

By assumption  $f_{b^*, a^*}(b | a^0)$  is symmetric around  $E(b | a^0)$ . This implies that<sup>7</sup>

$$b \in [2E(b | a^0) - \underline{b}, \bar{b}] \Rightarrow f_{b^*, a^*}(b | a^0) = 0$$

Therefore,  $F_{y^*, a^*}(y^0 | a^0)$  is equal to

$$\int_{\underline{b}}^{E(b|a^0)} F_{y^*, a^*, b^*}(y^0 | a^0, b) f_{b^*, a^*}(b | a^0) db \\ + \int_{E(b|a^0)}^{2E(b|a^0) - \underline{b}} F_{y^*, a^*, b^*}(y^0 | a^0, b) f_{b^*, a^*}(b | a^0) db$$

Now introduce a change of variable: replace  $b$  by  $E(b | a^0) + c$  to obtain:

$$\int_{\underline{b} - E(b|a^0)}^0 F_{y^*, a^*, b^*}(y^0 | a^0, E(b | a^0) + c) f_{b^*, a^*}(E(b | a^0) + c | a^0) dc \\ + \int_0^{E(b|a^0) - \underline{b}} F_{y^*, a^*, b^*}(y^0 | a^0, E(b | a^0) + c) f_{b^*, a^*}(E(b | a^0) + c | a^0) dc$$

Due to the symmetry of  $f_{b^*, a^*}(b | a^0)$  around  $E(b | a^0)$ , we have that this equals

$$\int_0^{E(b|a^0) - \underline{b}} [F_{y^*, a^*, b^*}(y^0 | a^0, E(b | a^0) + c) + F_{y^*, a^*, b^*}(y^0 | a^0, E(b | a^0) - c)] \cdot \\ f_{b^*, a^*}(E(b | a^0) + c | a^0) dc \quad (16)$$

Since  $\frac{\partial F_{y^*, a^*, b^*}(\mu | a^0, b)}{\partial b}$  is symmetric in  $b$  around  $E(b | a^0)$ ,

$$F_{y^*, a^*, b^*}(\mu | a^0, E(b | a^0) + c) + F_{y^*, a^*, b^*}(\mu | a^0, E(b | a^0) - c) \\ = 2F_{y^*, a^*, b^*}(\mu | a^0, E(b | a^0))$$

Therefore,

$$F_{y^*, a^*}(\mu | a^0) = \\ 2F_{y^*, a^*, b^*}(\mu | a^0, E(b | a^0)) \int_0^{E(b|a^0) - \underline{b}} f_{b^*, a^*}(E(b | a^0) + c | a_0) dc$$

In view of the symmetry of  $f_{b^*, a^*}(b | a^0)$  the integral in the last expression is equal to  $1/2$ , such that

$$F_{y^*, a^*}(\mu | a^0) = F_{y^*, a^*, b^*}(\mu | a^0, E(b | a^0)).$$

**Proof of corollary 20**

$$F_{y_t^*, y_{t-1}^*}(y_t^0 | y_{t-1}^0) = \Pr \{y_t^* \leq y_t^0 | y_{t-1}^* = y_{t-1}^0\} = F_e(y_t^0 - \beta y_{t-1}^0)$$

Using (8), we can re-write (7) as  $y_t + u_t = \beta y_{t-1}^* + e_t^y$ . We can then calculate the conditional distribution of this misspecified model as:

$$F_{y_t, y_{t-1}^*}(y_t^0 | y_{t-1}^* = y_{t-1}^0) = \Pr \{y_t \leq y_t^0 | y_{t-1}^* = y_{t-1}^0\} \\ = \Pr \{\beta y_{t-1}^* + e_t^y - u_t \leq y_t^0 | y_{t-1}^* = y_{t-1}^0\} \\ = \Pr \{e_t^y - u_t \leq y_t^0 - \beta y_{t-1}^0\}.$$

---

<sup>7</sup>We do the proof for the case where  $E(b|a_0) \leq (\underline{b} + \bar{b})/2$ . The proof for the case where the reverse inequality holds is similar.



## Appendix 2: The Equilibrium Distribution of the Intergenerational Model

In this appendix we derive the equilibrium distribution of incomes of children and parents, and the endowments of the child. The intergenerational model is described by equations (5) and (6), here repeated for convenience:

$$y_t^* = \beta y_{t-1}^* + \gamma b_t + e_t^y$$

$$b_t = \rho b_{t-1} + e_t^b$$

where  $e_t^y$  and  $e_t^b$  are iid with mean zero and variance  $\sigma_{ey}^2$  and  $\sigma_{eb}^2$ , respectively. We assume that  $0 < \beta < 1$  and  $0 < \rho < 1$ . We can then establish the following theorem:

**Theorem 21** *Solution to the Intergenerational Model:*

$$\begin{bmatrix} y_t^* \\ y_{t-1}^* \\ b_t \end{bmatrix} \sim N \left[ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Omega \right]$$

where  $\Omega =$

$$\begin{bmatrix} \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{1+\beta\rho}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{1}{1-\beta^2} \sigma_{ey}^2 & \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{\rho+\beta}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 & \frac{\gamma}{(1-\rho^2)(1-\beta\rho)} \sigma_{eb}^2 \\ \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{\rho+\beta}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 & \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{1+\beta\rho}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{1}{1-\beta^2} \sigma_{ey}^2 & \gamma\rho \frac{1}{1-\rho^2} \frac{1}{1-\beta\rho} \sigma_{eb}^2 \\ \frac{\gamma}{(1-\rho^2)(1-\beta\rho)} \sigma_{eb}^2 & \gamma\rho \frac{1}{1-\rho^2} \frac{1}{1-\beta\rho} \sigma_{eb}^2 & \frac{1}{1-\rho^2} \sigma_{eb}^2 \end{bmatrix}$$

**Proof :**

Note that

$$b_t = \sum_{i=0}^{\infty} \rho^i e_{t-i}^b \quad (17)$$

$$y_t^* = \gamma \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-i}^y \quad (18)$$

Consequently, the variables  $[ y_t^* \ y_{t-1}^* \ b_t ]$  are all infinite sums of iid variables, such that, by the central limit theorem, they will be jointly normally distributed. The mean of  $[ y_t^* \ y_{t-1}^* \ b_t ]$  is  $[ 0 \ 0 \ 0 ]$ . Next, we determine the elements of  $\Omega$ :

(1)

$$E (b_t)^2 = E \left[ (e_t^b + \rho e_{t-1}^b + \rho^2 e_{t-2}^b + \dots)^2 \right]$$

$$= (1 + \rho^2 + \rho^4 + \dots) \sigma_{eb}^2 = \frac{1}{1-\rho^2} \sigma_{eb}^2$$

(2)

$$y_t^* b_t = \gamma \left[ \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-i}^y \right] \left[ \sum_{i=0}^{\infty} \rho^i e_{t-i}^b \right]$$

$$E (y_t^* b_t) = \gamma \left[ (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \right. \\ \left. + \beta (\rho + \rho^3 + \rho^5 + \dots) \right. \\ \left. + \beta^2 (\rho^2 + \rho^4 + \rho^6 + \dots) + \dots \right] \sigma_{eb}^2$$

$$= \gamma \left[ \frac{1}{1-\rho^2} + \frac{\beta\rho}{1-\rho^2} + \frac{(\beta\rho)^2}{1-\rho^2} + \dots \right] \sigma_{eb}^2$$

$$= \frac{\gamma}{(1-\rho^2)(1-\beta\rho)} \sigma_{eb}^2$$

(3)

$$(y_t^*)^2 = \gamma^2 \left[ \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-i}^y \right]^2$$

$$E (y_t^*)^2 = \gamma^2 \cdot$$

$$\left[ (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \right. \\ \left. + \beta^2 (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \right. \\ \left. + \beta^4 (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \right. \\ \left. + \beta^6 (1 + \rho^2 + \rho^4 + \rho^6 + \dots) + \dots \right. \\ \left. + 2\beta (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \right. \\ \left. + 2\beta^2 (\rho^2 + \rho^4 + \rho^6 + \rho^8 + \dots) \right. \\ \left. + 2\beta^3 (\rho^3 + \rho^5 + \rho^7 + \rho^9 + \dots) + \dots \right. \\ \left. + 2\beta^3 (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \right. \\ \left. + 2\beta^4 (\rho^2 + \rho^4 + \rho^6 + \rho^8 + \dots) \right. \\ \left. + 2\beta^5 (\rho^3 + \rho^5 + \rho^7 + \rho^9 + \dots) + \dots \right. \\ \left. + \dots \right] \sigma_{eb}^2$$

$$+ [1 + \beta^2 + \beta^4 + \dots] \sigma_{ey}^2$$

$$E (y_t^*)^2 = \gamma^2 \frac{1}{1-\rho^2} \sigma_{eb}^2 \cdot$$

$$\left[ (1 + \beta^2 + \beta^4 + \dots) \right. \\ \left. + 2 (\beta\rho + \beta^2\rho^2 + \beta^3\rho^3 + \dots) \right. \\ \left. + 2 (\beta^3\rho + \beta^4\rho^2 + \beta^5\rho^3 + \dots) + \dots \right] + \frac{1}{1-\beta^2} \sigma_{ey}^2$$

$$E (y_t^*)^2 = \gamma^2 \frac{1}{1-\rho^2} \sigma_{eb}^2 \cdot$$

$$\begin{aligned}
& \left[ \frac{1}{1-\beta^2} + \frac{2\beta\rho}{1-\beta\rho} + \frac{2\beta^2(\beta\rho)}{1-\beta\rho} + \frac{2\beta^4(\beta\rho)}{1-\beta\rho} + \dots \right] + \frac{1}{1-\beta^2}\sigma_{ey}^2 \\
E(y_t^*)^2 &= \gamma^2 \frac{1}{1-\rho^2} \sigma_{eb}^2 \left[ \frac{1}{1-\beta^2} + \frac{2\beta\rho}{1-\beta\rho} (1 + \beta^2 + \beta^4 + \dots) \right] + \frac{1}{1-\beta^2}\sigma_{ey}^2 \\
E(y_t^*)^2 &= \gamma^2 \frac{1}{1-\rho^2} \sigma_{eb}^2 \frac{1}{1-\beta^2} \left[ 1 + \frac{2\beta\rho}{1-\beta\rho} \right] + \frac{1}{1-\beta^2}\sigma_{ey}^2 \\
E(y_t^*)^2 &= \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{1+\beta\rho}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{1}{1-\beta^2}\sigma_{ey}^2 \\
(4) \\
y_t^* y_{t-1}^* &= \gamma^2 \left[ \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-i}^y \right] \\
&\quad \cdot \left[ \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-1-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-1-i}^y \right] \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \cdot \\
&\quad [(\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^2 (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^4 (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) + \dots \\
&\quad + \beta (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \\
&\quad \quad + \beta^2 (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^3 (\rho^2 + \rho^4 + \rho^6 + \rho^8 + \dots) + \dots \\
&\quad + \beta^3 (1 + \rho^2 + \rho^4 + \rho^6 + \dots) \\
&\quad \quad + \beta^4 (\rho + \rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^5 (\rho^2 + \rho^4 + \rho^6 + \rho^8 + \dots) + \dots \\
&\quad + \beta (\rho^2 + \rho^4 + \rho^6 + \dots) \\
&\quad \quad + \beta^2 (\rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^3 (\rho^4 + \rho^6 + \rho^8 + \dots) + \dots \\
&\quad + \beta^3 (\rho^2 + \rho^4 + \rho^6 + \dots) \\
&\quad \quad + \beta^4 (\rho^3 + \rho^5 + \rho^7 + \dots) \\
&\quad \quad + \beta^5 (\rho^4 + \rho^6 + \rho^8 + \dots) + \dots] \sigma_{eb}^2 \\
&\quad + [\beta + \beta^3 + \beta^5 + \beta^7 + \dots] \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \sigma_{eb}^2 \frac{1}{1-\rho^2} \cdot \\
&\quad [\rho (1 + \beta^2 + \beta^4 + \dots) \\
&\quad \quad + (\beta + \beta^2 \rho + \beta^3 \rho^2 + \dots) \\
&\quad \quad + (\beta^3 + \beta^4 \rho + \beta^5 \rho^2 + \dots) + \dots \\
&\quad \quad + (\beta \rho^2 + \beta^2 \rho^3 + \beta^3 \rho^4 + \dots) \\
&\quad \quad + (\beta^3 \rho^2 + \beta^4 \rho^3 + \beta^5 \rho^4 + \dots) + \dots] + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \sigma_{eb}^2 \frac{1}{1-\rho^2} \cdot \\
&\quad \left[ \frac{\rho}{1-\beta^2} + \beta (1 + \beta\rho + (\beta\rho)^2 + \dots) + \beta^3 (1 + \beta\rho + (\beta\rho)^2 + \dots) \right. \\
&\quad \left. + \beta\rho^2 (1 + \beta\rho + (\beta\rho)^2 + \dots) + \beta^3 \rho^2 (1 + \beta\rho + (\beta\rho)^2 + \dots) \right]
\end{aligned}$$

$$\begin{aligned}
& + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \sigma_{eb}^2 \frac{1}{1-\rho^2} \left[ \frac{\rho}{1-\beta^2} + \frac{1}{1-\beta\rho} [(\beta + \beta^3 + \dots) + \rho^2 (\beta + \beta^3 + \dots)] \right] \\
& + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \sigma_{eb}^2 \frac{1}{1-\rho^2} \left[ \frac{\rho}{1-\beta^2} + \frac{1}{1-\beta\rho} \frac{\beta}{1-\beta^2} (1 + \rho^2) \right] + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \sigma_{eb}^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \rho + \beta \frac{1}{1-\beta\rho} (1 + \rho^2) \right] + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
E(y_t^* y_{t-1}^*) &= \gamma^2 \frac{1}{1-\rho^2} \frac{1}{1-\beta^2} \left[ \frac{\rho+\beta}{1-\beta\rho} \right] \sigma_{eb}^2 + \frac{\beta}{1-\beta^2} \sigma_{ey}^2 \\
(5) \\
y_{t-1}^* b_t &= \gamma \left[ \sum_{i=0}^{\infty} \beta^i \sum_{j=0}^{\infty} \rho^j e_{t-1-i-j}^b + \sum_{i=0}^{\infty} \beta^i e_{t-1-i}^y \right] \left[ \sum_{i=0}^{\infty} \rho^i e_{t-i}^b \right] \\
E(y_{t-1}^* b_t) &= \gamma \sigma_{eb}^2 \cdot \\
& \quad [(\rho + \rho^3 + \rho^5 + \dots) \\
& \quad + \beta(\rho^2 + \rho^4 + \rho^6 + \dots) \\
& \quad + \beta^2(\rho^3 + \rho^5 + \rho^7 + \dots) + \dots] \\
E(y_{t-1}^* b_t) &= \gamma \sigma_{eb}^2 \frac{1}{1-\rho^2} [\rho + \beta\rho^2 + \beta^2\rho^3 + \dots] \\
E(y_{t-1}^* b_t) &= \gamma \sigma_{eb}^2 \frac{1}{1-\rho^2} \rho [1 + \beta\rho + (\beta\rho)^2 + \dots] \\
E(y_{t-1}^* b_t) &= \gamma \rho \frac{1}{1-\rho^2} \frac{1}{1-\beta\rho} \sigma_{eb}^2 \\
(6) \\
E(y_{t-1}^*)^2 &= E(y_t^*)^2
\end{aligned}$$

Table 1: Omitted conditioning variables with positive (negative) covariance

	Low $a^0$	High $a^0$
$\frac{\partial F_{y^*, a^*, b^*}(y^0   a^0, b)}{\partial b} \leq 0$	+ (-)	- (+)
$\frac{\partial F_{y^*, a^*, b^*}(y^0   a^0, b)}{\partial b} \geq 0$	- (+)	+ (-)

Table 2:  $M(\alpha, y_{t-1})$  and  $M^e(\alpha, y_{t-1})$  in the case of measurement error in son's and father's income

	$M(0.1, y_{t-1}^0)$	$M^e(0.1, y_{t-1}^0)$ (error in son's income)	$M^e(0.1, y_{t-1}^0)$ (error in father's income)
$y_{t-1}^0   0.1$	.59	.65	.64
$y_{t-1}^0   0.2$	.71	.73	.74
$y_{t-1}^0   0.3$	.75	.76	.77
$y_{t-1}^0   0.4$	.76	.78	.78
$y_{t-1}^0   0.5$	.77	.78	.78
$y_{t-1}^0   0.6$	.76	.78	.78
$y_{t-1}^0   0.7$	.75	.76	.77
$y_{t-1}^0   0.8$	.71	.73	.74
$y_{t-1}^0   0.9$	.59	.65	.64

Table 3:  $M(\alpha, y_{t-1})$  and  $M^e(\alpha, y_{t-1})$  in the case of omitted ability

	$M^e(0.1, y_{t-1}^0)$ (1)	$M(0.1, y_{t-1}^0)$ $b = -0.37$ (2)	$M(0.1, y_{t-1}^0)$ $b = 0$ (3)	$M(0.1, y_{t-1}^0)$ $b = 0.37$ (4)
$y_{t-1}^0   0.1$	0.49	0.40	0.64	0.83
$y_{t-1}^0   0.2$	0.65	0.52	0.64	0.80
$y_{t-1}^0   0.3$	0.70	0.61	0.65	0.77
$y_{t-1}^0   0.4$	0.72	0.66	0.65	0.74
$y_{t-1}^0   0.5$	0.73	0.71	0.66	0.71
$y_{t-1}^0   0.6$	0.72	0.74	0.65	0.66
$y_{t-1}^0   0.7$	0.70	0.77	0.65	0.61
$y_{t-1}^0   0.8$	0.65	0.80	0.64	0.52
$y_{t-1}^0   0.9$	0.49	0.83	0.64	0.40

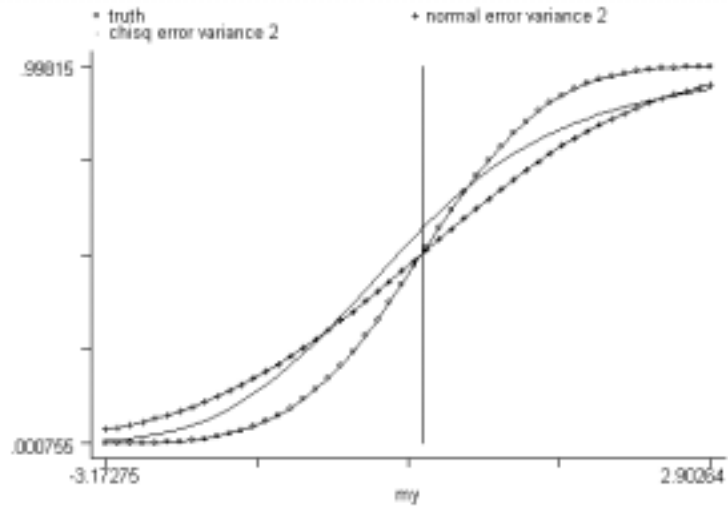


Figure 1: True Model Standard Normal with two misspecified models. The first uses errors from  $N(0,2)$  the second from a Chi-squared(1) adjusted to have mean zero.

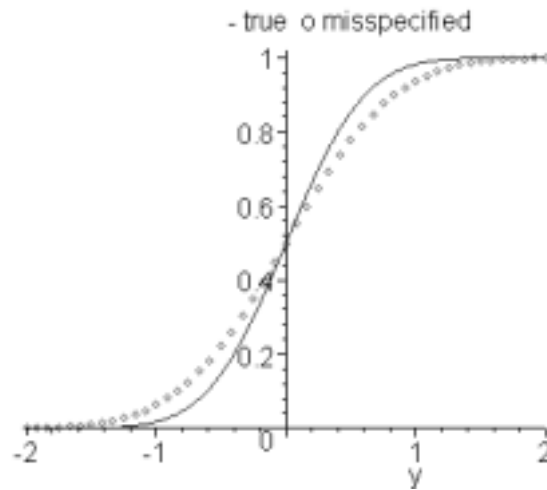


Figure 2: True Conditional Distribution and Distribution of  $Y$  conditional on the median value of  $X$  when  $Y$  is measured with error.

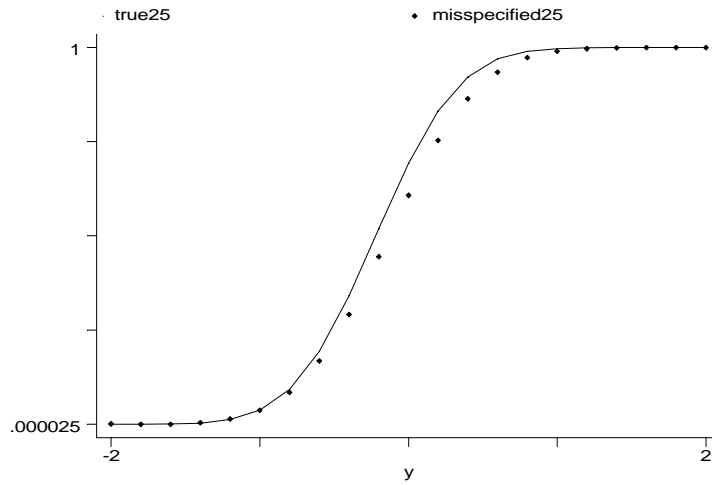


Figure 3: True Conditional Distribution and Distribution of Y conditional on the 25th percentile of X when X is measured with error.

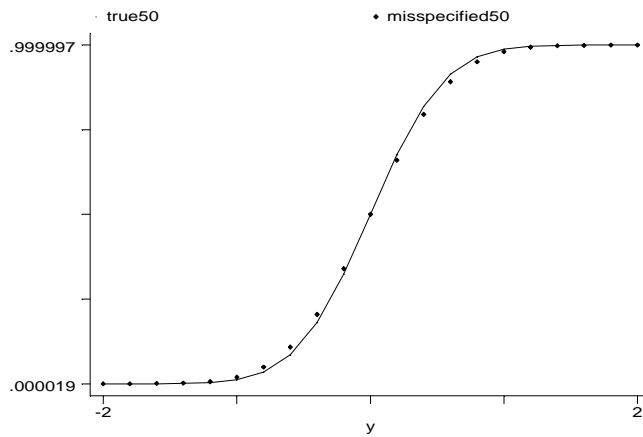


Figure 4: True Conditional Distribution and Distribution of Y conditional on the 50th percentile of X when X is measured with error.

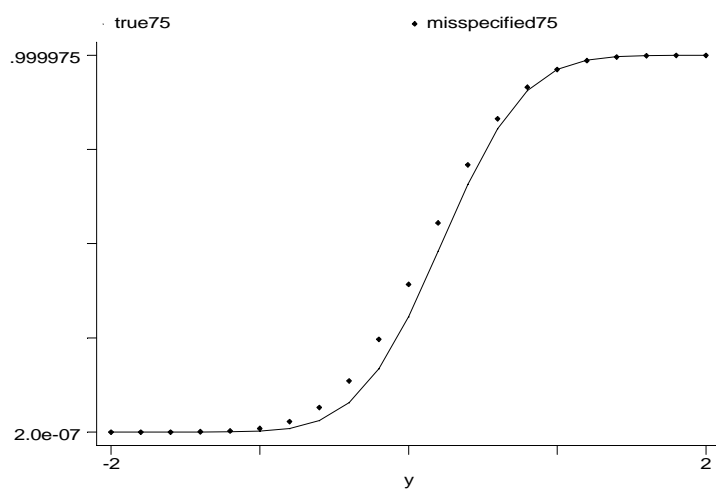


Figure 5: True Conditional Distribution and Distribution of Y conditional on the 75th percentile of X when X is measured with error.



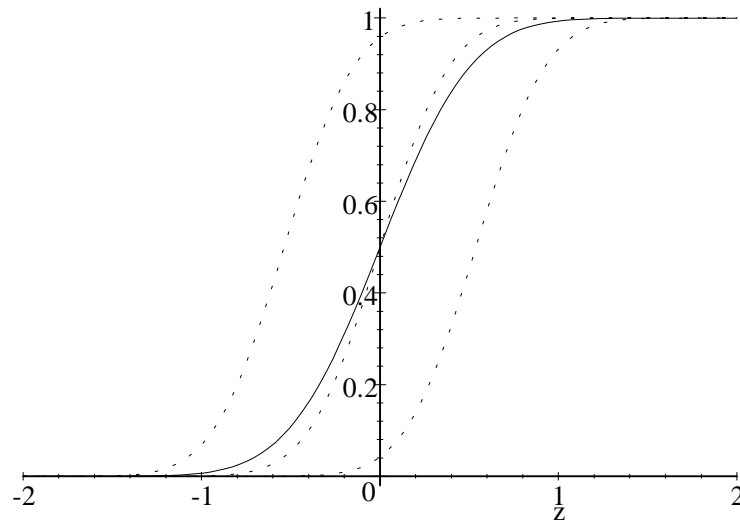


Figure 6: Omitted variables : The distribution of  $y_t$  given the mean value of  $y_{t-1}$ .  $F_{y_t^*, y_{t-1}^*, b}(y_t | E(y_{t-1}), b_{.025})$  (left dotted distribution),  $F_{y_t^*, y_{t-1}^*, b}(y_t | E(y_{t-1}), b_{.975})$  (right dotted distribution),  $F_{y_t^*, y_{t-1}^*, b}(y_t | E(y_{t-1}), b_{.5})$  (center dotted distribution) and  $F_{y_t^*, y_{t-1}^*}(y_t | E(y_{t-1}))$  (centre solid distyribution)

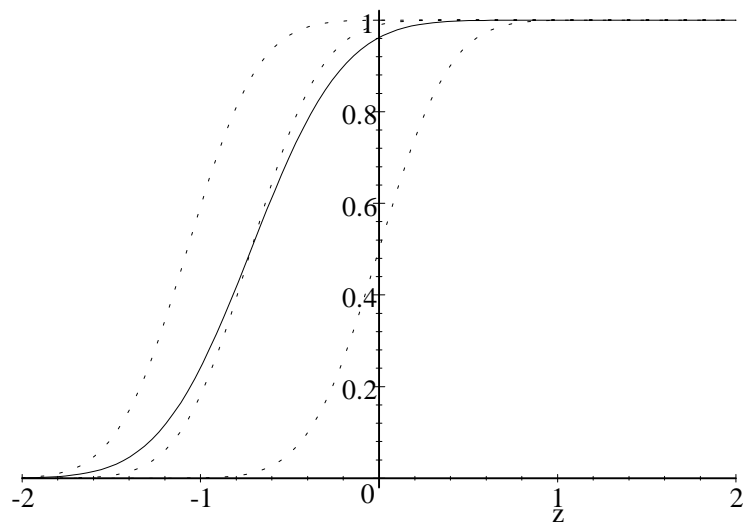


Figure 7: Omitted variables : The distribution of  $y_t$  given the 2.5th percentile of  $y_{t-1}$ .  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{.025}, b_{.025})$  (left dotted distribution),  $F_{y_t, y_{t-1}^*, b}(y_t | y_{.025}, b_{.975})$  (right dotted distribution),  $F_{y_t, y_{t-1}^*, b}(y_t | y_{.025}, E(b | y_{.025}))$  (center dotted distribution) and  $F_{y_t^*, y_{t-1}^*}(y_t | y_{.025})$  (centre solid distribution)

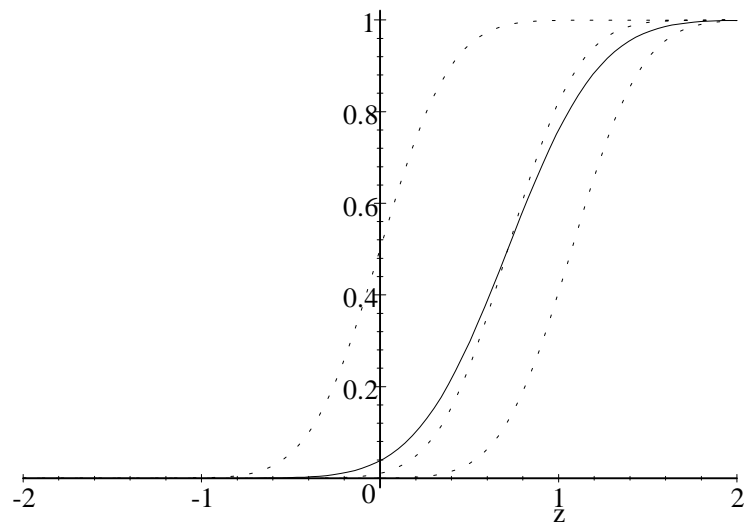


Figure 8: Omitted variables : The distribution of  $y_t$  given the 97.5th percentile of  $y_{t-1}$ .  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{.975}, b_{.025})$  (left dotted distribution),  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{.975}, b_{.975})$  (right dotted distribution),  $F_{y_t^*, y_{t-1}^*, b}(y_t | y_{.975}, E(b | y_{.975}))$  (center dotted distribution) and  $F_{y_t^*, y_{t-1}^*}(y_t | y_{.975})$  (centre solid distyribution)