

IZA DP No. 8143

Persistence Bias and Schooling Returns

Corrado Andini

April 2014

Persistence Bias and Schooling Returns

Corrado Andini
*Universidade da Madeira,
CEEApIA and IZA*

Discussion Paper No. 8143
April 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Persistence Bias and Schooling Returns^{*}

A well-established empirical literature suggests that individual wages are persistent. Several theoretical arguments support this empirical finding. Yet, the standard approach to the estimation of schooling returns does not account for this fact. This paper investigates the consequences of disregarding earnings persistence. In particular, it shows that the most commonly used static-model estimators of schooling coefficients are subject to an omitted-variable bias which can be named “persistence bias”.

JEL Classification: C23, I21, J31

Keywords: schooling, wages, dynamic panel-data models

Corresponding author:

Corrado Andini
Universidade da Madeira
Campus da Penteadá
9000-390 Funchal
Portugal
E-mail: andini@uma.pt

^{*} For valuable comments and suggestions on previous versions of this manuscript, the author would like to thank Monica Andini, Mário Centeno, Giuseppe Croce, Massimo Filippini, Francesco Franco, Vincenzo Galasso, Alberto Holly, Guido Maretto, Pedro Martins, Ilaria Maselli, Fabrizio Mazzonna, Álvaro Novo, Susana Peralta, Ana Balcão Reis, Stefano Staffolani, and other participants at presentations held in Rome (LUISS, Sep. 2013), Lugano (USI, Nov. 2013) and Lisbon (NovaSBE, Nov. 2013). Part of this paper has been written while the author was visiting the Economic Research Department at the Banco de Portugal, whose kind hospitality is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Since the publication of a seminal article by Griliches (1977), it is known that the ordinary least squares estimator of the schooling coefficient in a simple static wage-schooling model is biased. In particular, Griliches pointed out the existence of two types of biases, which are sometimes referred as the “Griliches’s biases”. The first, known as the “ability bias”, is an upward bias due to the correlation between individual unobserved ability and schooling.¹ The second, known as the “attenuation bias”, is a downward bias due to measurement errors in the schooling variable.

Attempts to cure or reduce the Griliches’s biases have been based on three main empirical approaches: i) extensions of the control set (to reduce the “importance” of the error term); ii) instrumental-variable estimation (to control for endogeneity); and iii) the use of longitudinal data (to control for individual unobserved heterogeneity). Of course, combinations of these approaches have also been adopted.

One striking feature of the existing literature is that the body of evidence is vast. This partly explains why it is difficult to make a definitive statement about the magnitude of the schooling coefficient, with and without correcting for the Griliches’s biases. However, one of the things that we know is that, as argued by Card (2001), instrumental-variable estimates of the schooling coefficient in static wage-schooling models are typically found to be bigger than least squares estimates,² and more imprecise. In this paper, we suggest that these estimates are both biased. The reason is that they generally rely on static specifications of the wage-schooling model which disregard earnings persistence.

Let us start with the least squares case. While there are hundreds of studies dealing with the Griliches’s biases, to the best of our knowledge, no research has been so far conducted to highlight another source of distortion for the least squares estimator of the schooling coefficient: the bias arising from a static wage-schooling model which disregards earnings persistence. We will refer to this omitted-variable bias as the “least squares persistence bias”.

¹Some authors, and Griliches himself, have questioned the existence of a necessarily positive correlation between schooling and ability by arguing that individuals endowed with higher ability have higher opportunity costs of attending school. If a negative correlation between schooling and ability is dominant, the least squares estimation of the schooling coefficient is subject to a downward ability bias.

²As suggested by Belzil (2007), this literature is known as the “instrumental-variable” or “experimental” literature. However, there exists another important branch of literature on wage-schooling models, known as the “structural” literature, in which the estimates of the schooling coefficient are typically found to be not only lower than the instrumental-variable estimates but also lower than the least squares estimates. In this paper, we investigate one possible explanation for this discrepancy in the estimates: the misspecification of the functional form of the wage-schooling model in the instrumental-variable literature. Indeed, as shown in Appendix, the standard model estimated in the instrumental-variable literature can be seen as a particular case of a more general dynamic wage-schooling model. For sake of clarification, our approach also differs from the structural approach because, while the latter is based on a dynamic discrete-choice model of schooling decisions ending up in a wage-schooling model where earnings persistence does not play any explicit role (or, at least, it is not estimated), we do not dynamically model the schooling decisions (likewise the instrumental-variable approach) but we see an explicit role for earnings persistence (unlike both the structural and the instrumental-variable approach) in the wage-schooling model. So, in a way, our approach is a dynamic instrumental-variable approach.

The first key issue in this paper is thus whether it is important or not to account for earnings persistence in a model for individual wages. Obviously, disregarding earnings persistence in wage-schooling models would not cause any problem if earnings persistence were not important in individual wage models. However, if earnings persistence were indeed important, then disregarding it would be problematic.

As a matter of fact, the empirical evidence on the persistent nature of earnings, both at micro and macro level, is already large. Indeed, it has already been reviewed, among others, by both Taylor (1999) and Guvenen (2009). The former has presented the macroeconomic evidence. The latter has discussed most of the existing microeconomic studies.

Focusing on the microeconomic evidence, which is particularly relevant for individual wage-schooling models, it is worth noting that the discussion about the persistence of individual wages is not new. Indeed, some of the first articles taking the dynamic aspects of individual earnings into account date back to the 1970s and the 1980s. Examples are given by Lillard and Willis (1978), MaCurdy (1982) as well as Abowd and Card (1989), among others. More recently, individual-level wage models taking the persistent nature of earnings into account have been proposed and estimated by Bell et al. (2002), Guiso et al. (2005), Cardoso and Portela (2009), Baltagi et al. (2009) and Hospido (2012), to cite a few.

With some exceptions, the above referred literature, and in particular the article by Guvenen (2009), models individual wages dynamically because the residuals of a static wage equation are seen as autocorrelated. From the labour demand perspective, one possible intuition is that, assuming real output is given by $Y = Af(L, K)$, the profit maximization of the employer leads to $w = y + a$ where $w = \ln W$ is the logarithm of the hourly wage, $y = \ln P \frac{\partial f(L, K)}{\partial L}$ is the logarithm of marginal labour productivity in money terms (which is a function of employee human capital and other employee and employer characteristics), and $a = \ln A$ is autoregressive of order one (see Appendix for details). The latter is an assumption usually made in the literature modelling total factor productivity (see King and Rebelo, 1999, among others), where the debate is not on the existence of persistence but rather on the magnitude of this persistence and how it changes over time. In short, total factor productivity shocks are assumed to be persistent. From the labour supply perspective, another possible intuition for autocorrelated residuals is that individuals are hit not only by transitory but also by permanent shocks (see Storesletten et al. 2004; among others), which can be thought as shocks to typically unobserved time-varying individual characteristics such as tastes.³

However, despite the existence of an already large body of empirical evidence on the persistence of individual wages, the incorporation of earnings persistence into human-capital and Mincerian-type models has been slow. One explanation for this fact is that it is uneasy to account for earnings persistence, endogeneity, individual unobserved heterogeneity and selection, all at the same time, even if the wage-schooling model is assumed to be linear. Nevertheless, the existing literature includes a few exceptions.

In particular, the importance of accounting for earnings persistence in wage-

³The individual labour supply can be seen as $w = f(L, R, \psi)$ where R is non-labour income and ψ is a parameter representing tastes.

schooling models has been repeatedly stressed by Andini (2007; 2009; 2010; 2013a; 2013b). For instance, Andini (2009; 2013a) has proposed a simple theoretical model to explain why past wages should play the role of additional explanatory variable in human-capital regressions. The intuition is that, in a world where bargaining matters, the past wage of an individual can affect his/her outside option and thus the bargained current wage. Analogously, Andini (2010; 2013b) has proposed an adjustment model between observed earnings and potential earnings (the latter being defined as the monetary value of the individual human-capital productivity) where the adjustment speed is allowed to be not perfect due to frictions in the labour market. In addition, Andini (2013b) has built a bridge between the literature on earnings dynamics (Guvenen, 2009) and the Mincerian literature, showing how to obtain a consistent GMM-SYS estimate of the schooling coefficient in a Mincerian wage equation when earnings persistence, endogeneity and individual unobserved heterogeneity are taken into account. Similarly, Semykina and Wooldridge (2013) have estimated a wage-schooling model accounting for earnings persistence and sample selection. Finally, Kripfganz and Schwarz (2013) have estimated a dynamic wage-schooling model using an econometric approach alternative to the GMM-SYS estimator.

Based on the above mentioned empirical micro evidence, this paper starts from the assumption that controlling for earnings persistence is potentially important in individual wage-schooling models. And, starting from this assumption, it elaborates on the consequences of disregarding the persistence of earnings in the least squares estimation of the schooling coefficient. In addition, this paper goes beyond the specific least-squares case by discussing the problems of other static-model estimators: i) those accounting for endogeneity and ii) those accounting for both individual unobserved heterogeneity and endogeneity. In particular, it will be argued that the use of the standard static instrumental-variable estimator does not solve the persistence-bias problem. Indeed, likewise the “least squares persistence bias” referred before, we will be able to provide an expression for an “instrumental-variable persistence bias”. Finally, it will be argued that using the Hausman-Taylor estimator, which accounts for both individual unobserved heterogeneity and endogeneity, is still problematic when the model disregards earnings persistence.

Specifically, this paper presents the following five novel findings. First, it provides an expression for the bias of the least squares estimator of the schooling coefficient in a simple wage-schooling model where earnings persistence is not accounted for. It is argued that the least squares estimator of the schooling coefficient is biased upward, and the bias is increasing with labour-market experience (or age) and the degree of earnings persistence. Second, data from the National Longitudinal Survey of Youth (NLSY) are used to show that the magnitude of the least squares persistence bias is non-negligible. Third, it is argued that the least squares persistence bias cannot be cured by increasing the control set. Fourth, an expression for the persistence bias of the standard instrumental-variable estimator of the schooling coefficient in a static wage-schooling model is provided. Finally, it is shown that disregarding earnings persistence is still problematic for the estimation of the schooling coefficient even if individual unobserved heterogeneity and endogeneity are taken into account. The case of the Hausman-Taylor estimator is considered. While the second and the third of the mentioned results are sample-specific, the others hold in general.

In short, the standard cures for the Griliches’s biases (based on extensions of the control set, treatments of endogeneity, and panel-data models) are unable to solve the persistence-bias problem related to the estimation of static wage-schooling models. Therefore, a large number of schooling-coefficient estimates, based on static models, is potentially subject to a persistence-bias critique.

The rest of the paper is organized as follows. Section 2 provides an expression for the persistence bias of the least squares estimator for the schooling coefficient. Section 3 investigates the magnitude of that bias using US data on young male workers. Section 4 analyzes whether the bias can be somehow reduced by extending the control set. Section 5 provides an expression for the persistence bias of the standard instrumental-variable estimator of the schooling coefficient. Section 6 explains why disregarding earnings persistence is still problematic even if individual unobserved heterogeneity and endogeneity are accounted for, using the Hausman-Taylor estimator. Section 7 explores the implications of the findings presented in the previous sections for the computation of schooling returns. Section 8 concludes.

2 Least squares persistence bias

This section provides an expression for the persistence bias of the least squares estimator of the schooling coefficient, under a set of simplifying hypotheses.

Let us consider a simple wage-schooling model. In particular, let us assume that the “true” model is as follows:

$$w_{i,s+z+1} = \alpha + \rho w_{i,s+z} + \beta s_i + u_{i,s+z+1} \quad (1)$$

where w is logarithm of gross hourly wage, $s \geq 1$ is schooling years, $z \geq 0$ is years of labour-market experience, and u is an error term.⁴ Hence, the “true” model is dynamic in the sense that past wages help to predict current wages.

As stressed in Section 1, there are at least three theoretical reasons why wage-schooling models should be handled as dynamic models: imperfect adjustment (Andini, 2010; 2013b), wage bargaining (Andini, 2009; 2013a) and autocorrelated residuals (Guvenen, 2009; Storesletten et al., 2004; among others). Details about each of these arguments are provided in Appendix.

In addition, let us assume that:

- (H1) $COV(s_i, u_{i,s+z+1}) = 0 \quad \forall i, s + z$
- (H2) $COV(w_{i,s+z}, u_{i,s+z+1}) = 0 \quad \forall i, s + z$
- (H3) $COV(u_{i,s+z}, u_{i,s+z+1}) = 0 \quad \forall i, s + z$
- (H4) $COV(u_{i,s+z}, u_{j,s+z}) = 0 \quad \forall i \neq j, s + z$
- (H5) $E(u_{i,s+z+1}) = 0 \quad \forall i, s + z$
- (H6) $V(u_{i,s+z+1}) = \theta^2 \quad \forall i, s + z$
- (H7) $V(s_i) = \sigma^2 \quad \forall i$
- (H8) $COV(s_i, \rho w_{i,s-1} + u_{i,s}) = 0 \quad \forall i, s$

⁴Following the standard Mincerian model, it is assumed that an individual starts working after leaving school. The first observed wage is observed in year s .

Assumption (H1) excludes the Griliches’s biases in order to focus on the persistence bias. Assumption (H2) is an additional condition required for the least squares estimator of model (1) to be consistent: it excludes the so-called “Nickell’s bias” (Nickell, 1981). Of course, both these assumptions are unlikely to hold. However, we will discuss the implications of removing them later on. First, we will use these simplifying assumptions to make the first point of this paper: the inconsistency of the least squares estimator of the schooling coefficient when the wage-schooling model does not take into account earnings persistence.

Assumptions from (H3) to (H7) are quite standard. Assumption (H8), instead, is not standard. It can be seen as an “initial condition”. One may think at $w_{i,s-1}$ as a reservation wage⁵ that every individual has in mind before leaving school, at time $s - 1$. Yet, this wage is not observed. Hence, at time s , the error term in model (1) will be given by $\rho w_{i,s-1} + u_{i,s}$. It may well be the case that this reservation wage is correlated with s_i as higher educated people are likely to have higher reservation wages. However, assumption (H8) excludes this possibility. The reason is simple and related to assumption (H1): at this stage, in order to focus on the least squares persistence bias, we exclude all sources of bias due to correlation between schooling and the error term in model (1). Again, we will discuss the implications of removing these simplifying assumptions later on.

Under the above hypotheses, a proof of the inconsistency of the least squares estimator applied to a simple static wage-schooling model is straightforward. In short, consider that the “true” model is (1) but earnings persistence is disregarded and the following static “false” model is estimated:

$$w_{i,s+z+1} = \alpha + \beta s_i + e_{i,s+z+1} \quad (2)$$

In (2), the actual nature of the error term is $e_{i,s+z+1} = \rho w_{i,s+z} + u_{i,s+z+1}$. However, this nature is disregarded and the assumption $\rho = 0$ is implicitly made. Then, it is easy to show that:

$$plim\beta_{OLS} = \beta + \rho \frac{COV(s_i, w_{i,s+z})}{V(s_i)} \quad (3)$$

If earnings persistence actually matters, i.e. $\rho \neq 0$, then the least squares estimator of β is biased, unless s_i and $w_{i,s+z}$ are uncorrelated. Intuitively, such orthogonality is unlikely to hold but, nevertheless, it is worth to provide a formal proof of the existence of correlation.

To begin with, let us remind that $V(s_i) = \sigma^2$. Hence, using assumption (H7), it can be shown that (details are provided in Andini, 2013c):

$$COV(s_i, w_{i,s+z}) = \beta\sigma^2(1 + \rho + \rho^2 + \dots + \rho^{z-1}) + \rho^z COV(s_i, w_{i,s}) \quad (4)$$

Since $COV(s_i, w_{i,s}) = \beta\sigma^2 + COV(s_i, \rho w_{i,s-1} + u_{i,s})$, we can use assumption (H8) to get the following result:

⁵The idea of a reservation wage is compatible with the presence of self-selection into the labour market. However, in this paper, we do not explicitly deal with this important issue. We just consider the estimation of a wage equation where earnings persistence, individual unobserved heterogeneity and endogeneity matter (see also footnote 6).

$$COV(s_i, w_{i,s+z}) = \beta\sigma^2(1 + \rho + \rho^2 + \dots + \rho^z) \quad (5)$$

The above expression proves that s_i and $w_{i,s+z}$ are, in general, correlated. Hence, using (3), it follows that:

$$plim\beta_{OLS} = \beta + \rho\beta \sum_{j=0}^z \rho^j \quad (6)$$

where $\rho\beta \sum_{j=0}^z \rho^j$ is the absolute “least squares persistence bias”. The conclusion is that the least squares estimator of the schooling coefficient in model (2) is biased upward if β and ρ are positive, with the bias being increasing in both ρ and z . Obviously, we can define the percent (or relative) bias as the ratio between the absolute bias and β . The latter is given by $\rho \sum_{j=0}^z \rho^j$, thus being independent of β .

As a matter of example, Figure 1 illustrates how the absolute persistence bias increases with z , assuming several degrees of earnings persistence and $\beta = 0.030$. Figure 2, instead, depicts the percent bias (times 100). The latter goes from a minimum of 30% ($z = 0$ and $\rho = 0.300$) to a maximum of 512% ($z = 7$ and $\rho = 0.900$). This means that, even for very low values of experience and earnings persistence, the percent bias is particularly severe. Of course, the lower the degree of earnings persistence is, the lower the percent bias is.

3 Is the least squares persistence bias worrisome?

It is interesting to discuss the magnitude of the persistence bias when estimating a simple static wage-schooling model with real data. Particularly, we find of interest to explore data from the National Longitudinal Survey of Youth (NLSY), a well-known dataset of US young workers, aged between 17 and 30, in which the persistence bias should be lower than in a standard dataset including older workers since the average experience (z) is lower.

The dataset, which contains observations on 545 males for the period of 1980-1987, has four main advantages: it is a balanced panel (which avoids a number of econometric issues with unbalanced panels), it is publically available (making replication easier), it has been already used in the literature⁶ (making comparison with earlier studies possible) and it has already been cleaned up, such that the schooling variable is actually time-invariant. The summary statistics of the variables and their meaning are presented in Table 1.

One thing to note is that the experience variable is calculated as $z = age - s - 6$, representing potential labour-market experience by definition. Hence, it is parallel to individual age since schooling is time-invariant. Moreover, it is also parallel to actual experience as the dataset does not include unemployment spells (there are no individuals with zero wages).

The estimation results, obtained using the least squares estimator, are presented in Table 2. Column 1 shows the estimates from model (1), the “true” dynamic

⁶To our knowledge, this dataset has been already used by Vella and Verbeek (1998), Wooldridge (2005) and Andini (2007; 2013a), among others.

one. The coefficient of schooling β is estimated at 0.034, with the degree of earnings persistence ρ estimated at 0.599. Column 2 provides the estimate of the schooling coefficient from the “false” static model (2), which does not control for earnings persistence. As expected, the estimate of the schooling coefficient is well above the “true” value of the coefficient. Indeed, the coefficient is estimated at 0.076. The difference between 0.076 and 0.034 can be seen as a proxy of the absolute persistence bias, under Section 2’s assumptions. Since the average experience (z) in the sample is 6.5 years and the degree of earnings persistence is roughly equal to 0.600, a 0.042 absolute bias is perfectly in line with our theoretical prediction in Section 2 (see Figure 1), and its magnitude is non-negligible (123%).

Of course, if Section 2’s assumptions do not hold, both the static- and the dynamic-model estimates are biased and the 0.042 difference between the two estimated schooling coefficients can be meaningless. In Section 5, we will take this point into account by trying to separate the persistence bias from other biases.

4 Does extending the control set help?

Columns 3 to 7 in Table 2 gradually extend the static model (2) to investigate whether the least squares persistence bias can be somehow reduced by increasing the control set, i.e. by improving the explanatory power of the static model (2) and searching for “substitutes” of the past wage.

For instance, column 3 proposes the classical Mincerian specification which controls for experience and its square. However, the coefficient of schooling does not decrease, thus indicating that experience (or age) is not a substitute for past wage. In contrast, the schooling coefficient increases to 0.102.

Columns from 4 to 7 add a number of individual specific characteristics, both time-varying and constant, which increase the explained variability of wages, though not as much as just controlling for past wage. The latter is easily verified by looking at the evolution of the R-squared coefficient. In particular, column 4 takes into account union membership, marital status, public-sector employment, race (whether the individual is Black or Hispanic; the excluded category is White) as well as presence of health disabilities. Column 5 adds information on the individual residence (whether the individual lives in the South, Northern Central or North East; the excluded category is North West). In addition, it controls for whether the individual lives in a rural area or not. Columns 6 and 7 add detailed information on industry and occupation, respectively. Hence, the estimates in column 7 are based on the full control set. The key finding is that no static specification is able to provide a coefficient of schooling close to the “true” one, estimated using model (1).

Table 3 performs some robustness checks by considering issues associated with i) the presence of year (or cohort) fixed effects, ii) the number of observations and iii) the existence of non-linearities.

To begin with, in column 2, year fixed effects are added to the full control set used in column 7 of Table 1. They are found to be not jointly significant (p-value 0.232). In addition, the R-squared coefficient does not significantly improve. Hence, likewise the experience variables, year effects cannot be seen as substitutes for past wage. At best, year effects can be seen as substitutes for experience variables them-

selves because, when we estimate model (2) without controlling for the experience variables, year effects turn out to be jointly significant (p-value 0.000). The intuition for this result is that time and experience variables are highly correlated (see the correlation matrix in Table 4), thus creating multicollinearity problems. It follows that, in order to obtain reliable inference, we should exclude either experience variables or year effects from the control set. Since the standard practice in the literature is to assume a Mincerian-type specification of the wage-schooling model, in order to keep the latter in the rest of this paper, we will continue keeping experience variables in the control set, thus excluding year effects.

Column 3 considers the possibility that a different number of observations (4,360 vs. 3,815) is at the root of the discrepancy between the estimates of the schooling coefficient. Hence, the static model is estimated by dropping the 1980 observations (which are missing in the dynamic model). Yet, the discrepancy does not vanish.

Finally, column 4 in Table 3 adds an interaction between schooling and experience to the full control set in order to allow for some degree of non-linearity in the wage-schooling model. Again, the key point of this section holds: no static specification provides a coefficient of schooling close to the “true” one, estimated using model (1). The coefficient of the interaction term is found significant at 5% level and negative but very close to zero (-0.003). This means that the return to schooling is slightly decreasing in labour market experience. As we will see, the latter is in contrast with our preferred schooling return vs. experience pattern. We will come back to this issue in Section 7.

Before concluding this section, it is worth stressing that, even if one is able to find a static specification of the wage-schooling model replicating the “true” schooling coefficient (using a good proxy for past wages), under the assumption that the “true” model is still the dynamic model (because earnings *are* persistent), the coefficient of schooling estimated using a static specification can only be interpreted as the return to schooling under the very unrealistic assumption that individuals never die (see Appendix for details). Hence, to recover the return to schooling in a more general setting, we still need an estimate of the degree of earnings persistence and therefore a dynamic approach.

5 Instrumental-variable persistence bias

So far, we have focused on the least squares estimator. Yet, as it is well known, the estimate of the schooling coefficient in model (1) based on the least squares estimator cannot be taken as a good proxy of the “true” value of the schooling parameter due to the correlation between errors and schooling (the Griliches’s biases) and/or between errors and lagged wage (the Nickell’s bias). Such correlation causes the least squares estimator of model (1) to be inconsistent. This section deals with this issue.

To fix the ideas, let us assume that the error term $u_{i,s+z+1}$ in model (1) would be better seen as the sum between individual-specific unobserved effects c_i , representing individual abilities or measurement errors in the schooling variable⁷, and a “well-

⁷If the reservation wage of an individual just depends on time-invariant characteristics of the individual, such as the schooling level, then it is time-invariant too and c_i can be assumed to

behaved” disturbance $v_{i,s+z+1}$. That is, let us assume that $u_{i,s+z+1} = c_i + v_{i,s+z+1}$ with:

- (H9) $COV(s_i, c_i) \neq 0 \quad \forall i$
- (H10) $COV(s_i, v_{i,s+z+1}) = 0 \quad \forall i, s+z$
- (H11) $COV(c_i, v_{i,s+z+1}) = 0 \quad \forall i, s+z$
- (H12) $COV(w_{i,s+z}, v_{i,s+z+1}) = 0 \quad \forall i, s+z$
- (H13) $COV(v_{i,s+z}, v_{i,s+z+1}) = 0 \quad \forall i, s+z$
- (H13) $COV(v_{i,s+z}, v_{j,s+z}) = 0 \quad \forall i \neq j, s+z$
- (H15) $E(v_{i,s+z+1}) = 0 \quad \forall i, s+z$
- (H16) $V(v_{i,s+z+1}) = \vartheta^2 \quad \forall i, s+z$

By introducing individual-specific unobserved effects correlated with schooling, we introduce several sources of bias for the least squares estimator applied to model (1). Indeed, assumption (H9) removes assumptions (H1) and (H8) and allows for the Griliches’s biases to exist. In addition, assumption (H9) removes assumption (H2) and allows for the Nickell’s bias to exist.

The literature has typically dealt with assumption (H9) using instrumental variables. However, while a big research effort has been oriented towards the search of the best instrumental variable, the presence of the past wage in model (1) has been generally neglected. Indeed, the standard practice has been to estimate the “false” static model, i.e. model (2), assuming $\rho = 0$ in the expression for the error term $e_{i,s+z+1} = \rho w_{i,s+z} + u_{i,s+z+1}$ where $u_{i,s+z+1} = c_i + v_{i,s+z+1}$. The key point of this section is precisely that the standard practice has been, in fact, incorrect because disregarding the past wage biases the instrumental-variable estimation of the schooling coefficient in model (2).

A simple proof of why a static instrumental-variable approach can be misleading is as follows. Let us suppose that a researcher worries about a possible correlation between $u_{i,s+z+1}$ and s_i because of c_i , but the role played by the past wage in model (1) is disregarded. The standard static instrumental-variable practice is to find a time-invariant external instrument g_i such that $COV(g_i, s_i) \neq 0$. In this case, it is easy to show that:

$$plim\beta_{IV} = \beta + \frac{COV(g_i, u_{i,s+z+1})}{COV(g_i, s_i)} + \rho \frac{COV(g_i, w_{i,s+z})}{COV(g_i, s_i)} \quad (7)$$

The conclusion is that, even if the researcher is able to find an instrument satisfying $COV(g_i, u_{i,s+z+1}) = 0$, i.e. the standard instrumental-variable assumption, the instrumental-variable estimator of β will still be inconsistent⁸ for $\rho \neq 0$ as $COV(g_i, s_i) \neq 0$ implies $COV(g_i, w_{i,s+z}) \neq 0$. This is trivial because $w_{i,s+z}$ is correlated with s_i and, thus, with g_i . The last term of the sum in expression (7) is the

capture this type of individual unobserved heterogeneity.

⁸Another source of bias for the instrumental-variable estimator in static models is the presence of heterogeneous returns to schooling, i.e. the case in which the schooling coefficient is not the same across individuals. There is a rapidly-growing body of literature on this topic with recent important contributions by Carneiro, Heckman and Vytlačil, among others. In this paper, we have not explored the intersection between heterogeneous returns and earnings persistence. However, the latter is an interesting topic for future research.

absolute “instrumental-variable persistence bias”.

This inconsistency result, based on a persistence-bias critique, appears to be of fundamental importance due to its implications for the standard static approach in the Mincerian and human-capital literature. In addition, it is also important for the experimental literature since, as stressed by Carneiro et al. (2006, p. 2), the instrumental-variable method “is the most commonly used method of estimating β . Valid social experiments or valid natural experiments can be interpreted as generating instrumental variables”. Yet, the autoregressive nature of wages is typically not taken into account in the experimental literature (see Angrist and Pischke, 2009, pp. 243-247).

6 Hausman-Taylor persistence bias

This section argues that disregarding earnings persistence is still problematic for the estimation of the schooling coefficient even if individual unobserved heterogeneity and endogeneity are taken into account. We will show that the persistence bias is a problem related to the estimation of a static wage-schooling model, regardless of whether this estimation is performed using an estimator which exploits the longitudinal structure of the dataset and takes both individual unobserved heterogeneity and endogeneity into account.

To make the point of this section, borrowing from Andini (2013b), we will first present a method to obtain consistent estimates of both the schooling coefficient and the degree of earnings persistence when individual unobserved heterogeneity, endogeneity and earnings persistence are taken into account. The method is based on the GMM-SYS estimator developed by Blundell and Bond (1998). Afterwards, we will focus on the distortion of the least squares estimator, which takes into account earnings persistence but disregards both individual unobserved heterogeneity and endogeneity. Finally, we will discuss the main point of this section by considering the Hausman-Taylor estimator, which takes into account individual unobserved heterogeneity and endogeneity but disregards earnings persistence.

6.1 GMM-SYS approach

Under the new assumptions made in Section 5, Andini (2013b) has shown that consistent⁹ estimates for β and ρ are obtained using the GMM-SYS estimator proposed by Blundell and Bond (1998), i.e. using the following system of equations:

$$\Delta w_{i,s+z+1} = \rho \Delta w_{i,s+z} + \Delta v_{i,s+z+1} \quad (8)$$

$$w_{i,s+z+1} = \alpha + \rho w_{i,s+z} + \beta s_i + c_i + v_{i,s+z+1} \quad (9)$$

⁹One limitation of the approach proposed by Andini (2013b) is that selection is not considered. A dynamic wage-schooling model where selection matters has been estimated by Semykina and Wooldridge (2013). Yet, in their approach, a non-zero correlation between the time-constant variables and time-invariant individual unobserved heterogeneity implies that the effect of time-constant observed variables, such as schooling, cannot be distinguished from that of the individual unobserved heterogeneity (Semykina and Wooldridge, 2013, p. 50).

and using $w_{i,s+z-1}$ and $\Delta w_{i,s+z-1}$ as instruments for (8) and (9), respectively.

Of course, the use of $\Delta w_{i,s+z-1}$ and further lags as instruments is the key assumption to identify the coefficient of schooling. This identification approach has three main advantages. First, it is theoretically well grounded. Indeed, it is based on a seminal article by Bils and Klenow (2000, p. 1163-1164) who have unambiguously showed that expected wage growth can positively affect schooling decisions. Since we typically have data on realized but not on expected wage growth, the identification argument is implicitly based on the hypothesis that realized wage growth is correlated with expected wage growth and thus with schooling. Of course, the correlation between realized wage growth and schooling is easily testable. Second, the above approach does not involve the choice of an external instrument since it is based on internal instruments. Third, the approach is easily testable. Indeed, all the orthogonality conditions imposed by the GMM-SYS estimator must pass the Hansen test, and the specific (additional) orthogonality conditions related to the level-equation (9) must pass the Difference-in-Hansen test.

In sum, the schooling-coefficient identification approach proposed above seems to be more flexible than the standard instrumental-variable approach, which, as stressed by Belzil (2007, p. 1090), typically uses a time-invariant external instrument. This is an advantage of the dynamic approach, which exploits the existence of internal instruments.

An additional requirement for consistent estimation is that the difference-equation instruments, i.e. $w_{i,s+z-1}$ and further lags, should not be weak. This may happen in presence of highly persistent variables. The latter is also an easily testable assumption. A test can be based on the estimation of an AR1 process with individual fixed effects for the variable in levels, again using the GMM-SYS estimator. Yet, a simpler test can be based on the least squares estimator, which typically overestimates the autoregressive coefficient (see Blundell and Bond, 2000). For instance, in our sample, using the least squares estimator, the autoregressive coefficient of the AR1 log-wage process (with constant term) is estimated at 0.626 with robust standard error of 0.025 and p-value equal to 0.000. Hence, the true autoregressive coefficient of the log-wage process is well below the critical value of 1.000. Moreover, the stationarity of the log-wage process is a sufficient condition for the required orthogonality between c_i and $\Delta w_{i,s+z-1}$ or further lags.

Using the full control set, the GMM-SYS estimator provides an estimate of the degree of earnings persistence ρ equal to 0.174 and an estimate of the schooling coefficient β equal to 0.102, both significant at 1% level. All the standard specification tests are passed. The main point is that we find evidence of earnings persistence, despite its degree appears to be quite low. Nevertheless, as we shall see in both the rest of this section and the next one, even a very low degree of earnings persistence can be highly problematic, when disregarded.

6.2 Bias in dynamic least squares models

Taking the above estimates as the “true” values of the corresponding parameters, it is interesting to discuss the biases implied by alternative estimators or models, with special attention to the coefficient of schooling.

The first thing to note is that Andini (2013b) has already investigated the consequences for the least squares estimator of introducing assumption (H9). In particular, using Belgian data, the author has pointed to an upward-biased estimate of the degree of earnings persistence and to a downward-biased estimate of the schooling coefficient.

Estimation with NLSY data in Table 5 confirms the above view. Column 1 reports the least squares estimates of model (1) with no controls. Column 2 is based on the full control set. The finding is that there is no big difference in the estimates of both β and ρ between column 1 and column 2. However, once individual unobserved heterogeneity and endogeneity are taken into account using the GMM-SYS estimator, the picture is different. Indeed, column 3 shows that the least squares estimator, used in column 2 and column 1, seems to overestimate the degree of earnings persistence and to underestimate the schooling coefficient.

Clearly, the problem with the least squares approach to model (1) is that it accounts neither for individual unobserved heterogeneity nor for endogeneity. Formally, this means that the implicit assumption $E(c_i + v_{i,s+z+1} | s_i, w_{i,s+z}) = 0$ does not hold.

6.3 Persistence bias in static panel data models

Yet, the key point in this section is not about the failure of dynamic least squares models. The key point here is to highlight how misleading can be the static-model estimation of the schooling coefficient, even when the control set is large and when both individual unobserved heterogeneity and endogeneity are taken into account. To this end, Table 6 presents some additional evidence comparing the “true” estimate of the schooling coefficient based on the GMM-SYS estimator, reported in column 4, with an estimate based on a well-known instrumental-variable estimator for static panel data models. In particular, we consider an estimator which is typically used when time-invariant variables, such as schooling, are included in the explanatory set: the Hausman-Taylor estimator. As a benchmark, we also report estimates of the schooling coefficient based on two different estimators for static panel data models: the random effects estimator and the Mundlak estimator.

The random effects estimator, used in column 1 of Table 6, exploits the longitudinal nature of the dataset by controlling for individual unobserved effects under the assumption that they are uncorrelated with schooling and other explanatory variables. The Mundlak estimator, used in column 2, assumes that the vector of individual unobserved effects can be seen as a linear function of the matrix of the mean values of the time-varying explanatory variables plus a vector of residual unobserved individual effects. This approach assumes that controlling for the above matrix in the random effects model is enough to break any correlation between the residual unobserved individual effects and the explanatory variables, including schooling. Finally, the Hausman-Taylor estimator, used in column 3, fully takes into account that schooling and other explanatory variables (but not all) can be correlated with individual unobserved effects, thus being endogenous. Hence, the Hausman-Taylor estimator takes both individual unobserved heterogeneity and endogeneity into account, although it disregards earnings persistence.

In all the columns of Table 6, the control set used is the full one. In particular, in the Hausman-Taylor estimation, the health status is taken as time-varying exogenous, the race indicator variables are taken as time-invariant exogenous, schooling is taken as time-invariant endogenous, and all the other variables in the full control set are taken as time-varying endogenous. The identification is based on the standard Hausman-Taylor approach. In particular, the mean value of the health status is used as instrument for schooling. Clearly, this identification approach can be criticized but, to our knowledge, it is the best one available with our dataset.

Focusing on the Hausman-Taylor estimation, the conclusion seems to be that again, likewise the classical instrumental-variable case in Section 5, disregarding earnings persistence can be problematic. Indeed, the coefficient of schooling based on the Hausman-Taylor estimator (0.220) more than doubles the “true” one (0.102). This is the key result of the comparison between column 3 and column 4 in Table 6. The intuition is straightforward: the implicit Hausman-Taylor assumption $E(\rho w_{i,s+z} + v_{i,s+z+1} | s_i, \dots, c_i) = 0$ does not hold.

For comparison, the schooling coefficients estimated using the Mundlak approach or the random effects approach seem to be downward biased, with the bias being relatively lower in the latter case.

7 Computation of schooling returns

One very important implication of the findings presented in the previous sections is that the return to schooling cannot be consistently estimated using a static wage-schooling model. The latter point is trivial: since the schooling return and the coefficient of schooling are the same thing in the static model, and the estimate of the schooling coefficient is, in general, biased in the static model as we have seen, it follows that the estimate of the schooling return is, in general, biased too. Yet, a prediction about the direction of the bias is difficult. It depends on the number and the type of the covariates in the wage-schooling model, on whether individual unobserved heterogeneity is taken into account, on whether endogeneity is taken into account and, in the case it is, on the number and the type of the instrumental variables used.

This raises the question of how to compute the schooling return after consistently estimating the coefficient of schooling and the degree of earnings persistence in the dynamic model. The first thing to note is that, in the dynamic approach, the return to schooling does not generally coincide with the coefficient of schooling. The second thing is that the schooling return is not independent of labour-market experience (or age) as in the static model. In particular, Andini (2013b) has shown that the schooling return is a function of experience, say $r(z)$. Its exact expression is $r(z) = \beta \sum_{j=0}^z \rho^j$.

The above expression implies that the schooling return is equal to β at the start of the working life ($z = 0$), and then it converges to the asymptotic value $\frac{\beta}{1-\rho}$. The more earnings are persistent, the slower the convergence is. For positive values of ρ , the return to schooling is lower at the start of the working life. The intuition is that, since one source of earnings persistence is related to labour-market imperfections preventing the instantaneous adjustment between wages and human-

capital productivity, the employer does not fully pay the education premium to the employee at the start of the working life because the former does not have full information about the actual productivity of the latter. However, with time, the employer gets informed about the employee and progressively adjusts the premium to its steady-state value.¹⁰

As a matter of example, Figure 3 plots the function $r(z)$ from labour-market entry ($z = 0$) to the seventh year of the working life ($z = 7$). The values for β and ρ are those reported in column 3 of Table 5 (the GMM-SYS estimates). The schooling return goes from an initial 10% at the start of the working life to a stable 12.5% after two years of working life. This is consistent with Farber (1994) who also uses NLSY data and argues that two thirds of new full-time jobs end within two years. Thus, after two years, the employer-employee match is stable and asymmetric information problems are less distorting. The result is also consistent with Andini (2013b) who finds an entry return of 6.3% for Belgian male workers, which goes up to roughly 9.3% after two years.

One implication of the above result is that the wage-schooling profiles are not parallel across experience groups during the first years of the working life, as shown in Figure 4. Regardless of the schooling level, the slope of the wage-schooling profile, i.e. the return to schooling, is higher for more experienced workers.

An additional point worth making is that, if we allow for the return to schooling to be dependent on experience in a static Hausman-Taylor model, i.e. we add an schooling-experience interaction term to the control set, we find that the coefficient of the interaction term is not statistically significant (p-value 0.547) and very close to zero from the left (-0.0009), likewise the least squares case in Section 4 (though in the least squares case the coefficient is found significant). So, the static model, even when it accounts for individual unobserved heterogeneity and endogeneity, is unable to capture the fact that, during the first years of the working life, the wage-schooling profiles are not parallel across experience groups and that the schooling return is increasing in experience (the interaction coefficient is found not significant and slightly negative, while it should have been significant and positive in order to mimic the results of the dynamic model).

Two additional things are worth mentioning about Figure 3. First, all the estimations based on static models fail to capture the initial increasing pattern of the schooling return. This is because the static approach generally assumes that the return to schooling is independent of labour market experience. Even if the use of a schooling-experience interaction term allows for an experience-dependent schooling return, the static approach is still biased because a relevant explanatory variable, the past wage, is arbitrarily excluded from the control set. Second, all static-model estimations fail to capture the steady-state value of the schooling return by far. The

¹⁰A common view is that the coefficient of schooling in a static model is the “long-run” return to schooling, while the schooling coefficient in a dynamic model is the “short-run” return. According to this view, these coefficients are naturally different and, in particular, the “short-run” return is naturally lower than the “long-run” return. Hence, both the static approach and the dynamic one can be used, depending on the type of return we are interested in. Unfortunately, this argument is wrong. This is for two reasons. First, if earnings are persistent, the static model provides a biased estimate of the coefficient of schooling. Second, the “short-run” return is, in fact, the return at labour-market entry $r(0) = \beta$. Hence, the expression “short-run” is misleading.

conclusion is that disregarding earnings persistence is problematic even when individual unobserved heterogeneity and endogeneity are taken into account (the case of the Hausman-Taylor estimator), and even when the degree of earnings persistence is very low (Figure 3 considers the case of a 0.174 degree of earnings persistence).

The relevance of the above implications for the literature on schooling returns is straightforward.

8 Conclusions

There are at least three intuitive reasons why wage-schooling models should be handled as dynamic models: i) individual human-capital productivity and wages may not adjust instantaneously due to frictions in the labour market (Andini, 2010; 2013b); ii) past wages may affect the outside option of an individual in a simple bargaining model over wages and productivity (Andini, 2009; 2013a); iii) the residuals of the wage equation may show some degree of persistence (Guvenen, 2009; Storesletten et al., 2004; among many others). Of course, combinations of these explanations enrich the set of possibilities.

Despite the above theoretical arguments and an already large body of evidence supporting the dynamic behaviour of individual wages, the existing human-capital literature has not paid sufficient attention to the dynamic nature of wage-schooling models. Indeed, while examples of estimated static wage-schooling models are abundant, examples of estimated dynamic wage-schooling models can be counted on the fingers of one hand.

This pattern of the human-capital literature, however, should not be surprising. The initial theoretical wage-schooling models put forward by the fathers of modern education economics (Becker, Ben-Porath and Mincer, to cite a few) were particularly clever and their predictions have inspired a large body of static model evidence. In addition, longitudinal datasets including information on individual characteristics have not been easily accessible for several decades, making dynamic micro-level empirical analyses not executable. Fortunately, at least with respect to the latter aspect, today's reality is different. Longitudinal datasets are abundant (sometimes freely available) and the issue raised in this paper can now receive the appropriate consideration from the research community. Whether this will happen or not is still an open question.

Starting from the above motivation, this paper has investigated the consequences of disregarding earnings persistence and estimating a static wage-schooling model. Five main results have been presented. First, the least squares estimator of the schooling coefficient has been shown to be biased upward, with a bias increasing in labour-market experience (age) and the degree of earnings persistence. Second, the least squares persistence bias has been found to be non-negligible in NLSY data. Third, the least squares persistence bias has been found to be non-curable by increasing the control set. Fourth, the standard static instrumental-variable approach has been shown to be inconsistent. Finally, disregarding earnings persistence has been argued to be still problematic even when the estimator used accounts for individual unobserved heterogeneity and endogeneity. The case of the Hausman-Taylor estimator has been discussed. Of course, the second and the third of our findings are

specific to the sample used. However, the other results hold in general.

To conclude, the main implication of the above mentioned findings for the human-capital literature is that the return to the schooling investment is unlikely to be consistently estimated using a static wage-schooling model which disregards earnings persistence. Yet, this conclusion is conditional on the hypotheses made in the paper and subject to a number of caveats. Indeed, there are several assumptions behind our dynamic wage-schooling model which can be removed and a number of possible extensions of the setting studied.

To begin with, we have assumed that the schooling variable is time-invariant and measured in years. One may think to a model where the schooling variable is time-varying (for instance, in the case in which the individuals are working students). Another possibility is to measure education attainment in levels, degrees or qualifications (say primary, secondary and tertiary education) as in Budría and Pereira (2005), among others. In addition, it is possible to investigate the presence of sheepskin effects by transforming schooling years into indicator variables as in Hungerford and Solon (1987), among many others. A further possibility is to use a polynomial function of the schooling variable rather than just a level variable as suggested by Trostel (2005), among others.

Second, we have assumed that the schooling coefficient is constant across individuals and non-random. In contrast, it can be assumed that the schooling coefficient is individual-specific as in Card (1995), among others. Another possibility is to model the schooling coefficient and all the other covariate coefficients as random variables. This setting naturally leads to a quantile-regression specification of the wage equation where one can measure the impact of schooling not only on the mean but also on the shape of the conditional wage distribution, i.e. at several quantiles, as in the static model first estimated by Buchinsky (1994) and then followed by many others. To this respect, it is worth noting that there are a couple of recent articles dealing with the estimation of dynamic quantile regression models with fixed effects (Galvão, 2011; Galvão and Montes-Rojas, 2010).

Third, we have disregarded the selection issue despite there are at least two forms of selection that seem worth studying. The first has to do with the participation in the labour market. We have explored a sample of individuals which contains neither unemployment spells nor zero wages. Yet, different samples may include such information. In this case, it is possible to model a participation equation on the lines of Semykina and Wooldridge (2013). The second form has to do with the schooling choice which can be dynamically modelled as in the structural approach reviewed by Belzil (2007).

Fourth, our estimated dynamic wage-schooling model controls neither for firm characteristics nor for firm fixed effects. Estimation based on matched employer-employee datasets, as in Carneiro et al. (2012) for example, may shed new light on the sources of wage persistence.

Finally, we have assumed that schooling and experience are separable, that the wage-schooling model is linear in parameters and that it can be estimated using parametric estimation techniques. One can go beyond each of these assumptions by imposing non-separability between schooling and experience, by estimating a non-linear model or even by estimating a dynamic wage-schooling model non-parametrically.

References

Abowd, J., Card, D. (1989) “On the covariance structure of earnings and hours changes”. *Econometrica*, 57(2): 411-445.

Andini, C. (2007) “Returns to education and wage equations: a dynamic approach”. *Applied Economics Letters*, 14(8): 577-579.

Andini, C. (2009) “Wage bargaining and the (dynamic) Mincer equation”. *Economics Bulletin*, 29(3): 1846-1853.

Andini, C. (2010) “A dynamic Mincer equation with an application to Portuguese data”. *Applied Economics*, 42(16): 2091-2098.

Andini, C. (2013a) “How well does a dynamic Mincer equation fit NLSY data? Evidence based on a simple wage-bargaining model”. *Empirical Economics*, 44(3): 1519-1543.

Andini, C. (2013b) “Earnings persistence and schooling returns”. *Economics Letters*, 118(3): 482-484.

Andini, C. (2013c) “Persistence bias and the wage-schooling model”. IZA discussion paper n. 7186, Institute for the Study of Labor, January.

Angrist, J., Pischke, J. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press.

Baltagi, B., Blien, U., Wolf, K. (2009) “New evidence on the dynamic wage curve for Western Germany: 1980-2004”. *Labour Economics*, 16(1): 47-51.

Bell, B., Nickell, S., Quintini, G. (2002) “Wage equations, wage curves and all that”. *Labour Economics*, 9(3): 341-360.

Belzil, C. (2007) “The return to schooling in structural dynamic models: a survey”. *European Economic Review*, 51(5): 1059-1105.

Bils, M., Klenow, P. (2000) “Does schooling cause growth?”. *American Economic Review*, 90(5): 1160-1183.

Blundell, R.W., Bond, S.R. (1998) “Initial conditions and moment restrictions in dynamic panel data models”. *Journal of Econometrics*, 87(1): 115-143.

Blundell, R.W., Bond, S.R. (2000) “GMM estimation with persistent panel data: an application to production functions”. *Econometric Reviews*, 19(3): 321-340.

- Buchinsky, M. (1994) “Changes in the U.S. wage structure 1963-1987: application of quantile regression”, *Econometrica*, 62(2): 405-458.
- Budría, S., Pereira, P.T. (2005) “Educational qualifications and wage inequality: evidence for Europe”. IZA discussion paper n. 1763, Institute for the Study of Labor, September.
- Card, D. (1995) “Earnings, schooling, and ability revisited”. In: Polachek, S. (ed.) *Research in Labor Economics*, Greenwich: JAI Press.
- Card, D. (2001) “Estimating the return to schooling: progress on some persistent econometric problems”. *Econometrica*, 69(5): 1127-1160.
- Cardoso, A., Portela, M. (2009) “Micro foundations for wage flexibility: wage insurance at the firm level”. *Scandinavian Journal of Economics*, 111(1): 29-50.
- Carneiro, A., Guimarães, P., Portugal, P. (2012) “Real wages and the business cycle: accounting for worker, firm, and job title heterogeneity”. *American Economic Journal: Macroeconomics*, 4(2): 133-152.
- Carneiro, P., Heckman, J., Vytlačil, E. (2006) “Estimating marginal and average returns to education”. Unpublished manuscript.
- Farber, H. (1994) “The analysis of interfirm worker mobility”, *Journal of Labor Economics*, 12(4): 554-593.
- Galvão, A.F. (2011) “Quantile regression for dynamic panel data with fixed effects”. *Journal of Econometrics*, 164(1): 142-157.
- Galvão, A.F, Montes-Rojas, G.V. (2010) “Penalized quantile regression for dynamic panel data”. *Journal of Statistical Planning and Inference*, 140(11): 3476-3497.
- Griliches, Z. (1977) “Estimating the returns to schooling: some econometric problems”. *Econometrica*, 45(1): 1-22.
- Guiso, L., Pistaferri, L., Schivardi, F. (2005) “Insurance within the firm”. *Journal of Political Economy*, 113(5): 1054-1087.
- Guvenen, F. (2009) “An empirical investigation of labor income processes”. *Review of Economic Dynamics*, 12(1): 58-79.
- Hospido, L. (2012) “Modelling heterogeneity and dynamics in the volatility of individual wages”. *Journal of Applied Econometrics*, 27(3): 386-414.

Hungerford, T., Solon, G. (1987) “Sheepskin effects in the returns to education”. *Review of Economics and Statistics*, 69(1): 175-177.

King, R.G., Rebelo, S.T. (1999) “Resuscitating real business cycles”. In: Taylor J.B., Woodford M. (Eds.) *Handbook of Macroeconomics*, New York: North Holland.

Kripfganz, S., Schwarz, C. (2013) “Estimation of linear dynamic panel data models with time-invariant regressors”. DB discussion paper n. 25/2013, Deutsche Bundesbank, May (last update: 16 December 2013).

Lillard, L., Willis, R. (1978) “Dynamic aspects of earnings mobility”. *Econometrica*, 46(5): 985-1012.

MaCurdy, T. (1982) “The use of time-series processes to model the error structure of earnings in longitudinal data analysis”. *Journal of Econometrics*, 18(1): 83-114.

Nickell, S. (1981) “Biases in dynamic models with fixed effects”. *Econometrica*, 49(6): 1417-1426.

Semykina, A., Wooldridge, J.M. (2013) “Estimation of dynamic panel data models with sample selection”. *Journal of Applied Econometrics*, 28(1): 47-61.

Storesletten, K., Telmer, C., Yaron, A. (2004) “Consumption and risk sharing over the life cycle”. *Journal of Monetary Economics*, 51(3): 609-633.

Taylor, J.B. (1999) “Staggered price and wage setting in macroeconomics”. In: Taylor J.B., Woodford M. (Eds.) *Handbook of Macroeconomics*, New York: North Holland.

Trostel, P.A. (2005) “Nonlinearity in the return to education”. *Journal of Applied Economics*, 8(1): 191-202.

Vella, F., Verbeek, M. (1998) “Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men”. *Journal of Applied Econometrics*, 13(2): 163-183.

Webbink, D., Hartog, J. (2004) “Can students predict starting salaries? Yes!”. *Economics of Education Review*, 23(2): 103-113.

Wooldridge, J.M. (2005) “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved effects”. *Journal of Applied Econometrics*, 20(1): 39-54.

Appendix. A general wage-schooling model

Suppose individual log-productivity ($y_{i,s+z+1}$) is a linear function of time-invariant observed schooling years (s_i), time-invariant unobserved abilities (q_i), which are allowed to be correlated with schooling years, and other time-varying variables ($x_{i,s+z+1}$). In short, we have:

$$y_{i,s+z+1} = \pi q_i + \lambda s_i + \gamma x_{i,s+z+1} \quad (\text{A1})$$

The standard human-capital theory suggests that:

$$w_{i,s+z+1} = y_{i,s+z+1} + v_{i,s+z+1} \quad (\text{A2})$$

(Standard model, implicit version)

or alternatively:

$$w_{i,s+z+1} = \pi q_i + \lambda s_i + \gamma x_{i,s+z+1} + v_{i,s+z+1} \quad (\text{A3})$$

(Standard model, explicit version)

where the error $v_{i,s+z+1}$ is assumed to be i.i.d. with zero mean and constant variance.

Define $\theta \in [0, 1]$. It can be shown that the standard model (A2) (or (A3)) is a particular case of each of the following three models where $\theta = 1$:

$$w_{i,s+z+1} - w_{i,s+z} = \theta(y_{i,s+z+1} - w_{i,s+z}) + v_{i,s+z+1} \quad (\text{A4})$$

(Adjustment model)

$$w_{i,s+z+1} = (1 - \theta)w_{i,s+z} + \theta y_{i,s+z+1} + v_{i,s+z+1} \quad (\text{A5})$$

(Wage bargaining model)

$$w_{i,s+z+1} = y_{i,s+z+1} + a_{i,s+z+1} \quad \text{where} \quad a_{i,s+z+1} = (1 - \theta)a_{i,s+z} + v_{i,s+z+1} \quad (\text{A6})$$

(Autocorrelated disturbances model)

In (A4), θ represents the speed of adjustment between (log) wages and (log) human-capital productivity. In (A5), θ can be seen as the bargaining power of the employee. In (A6), $1 - \theta$ is the degree of persistence in the model (log) disturbances representing shocks in total factor productivity (demand-side view) or tastes (supply-side view). For a more detailed discussion about (A4), see Andini (2010; 2013b). For a discussion about (A5), see Andini (2009; 2013a). For a discussion about (A6), see Guvenen (2009) and Storesletten et al. (2004), among others.

The above three models can be all written as one single model, by appropriately re-labelling parameters. Specifically, let $\rho = 1 - \theta$, $\beta = \theta\lambda$, $c_i = \theta\pi q_i$ and suppose $\gamma = 0$ to simplify the exposition. It follows that:

$$w_{i,s+z+1} = \rho w_{i,s+z} + \beta s_i + c_i + v_{i,s+z+1} \quad (\text{A7})$$

(General model, dynamic version)

This is the general wage-schooling model referred in the title of this Appendix, though it can be made even more general in a number of directions. The coefficient of schooling in the static model (λ) only coincides with that of the dynamic model ($\beta = \theta\lambda$) in a very special case ($\theta = 1$). In general ($\theta < 1$), it is higher ($\lambda > \beta$).

Using backward substitution, we can write model (A7) as follows:

$$w_{i,s+z} = \rho^{z+1} w_{i,s-1} + \beta \left(\sum_{j=0}^z \rho^j \right) s_i + \left(\sum_{j=0}^z \rho^j \right) c_i + \sum_{j=0}^z \rho^j v_{i,s+z-j} \quad (\text{A8})$$

(General model, static version)

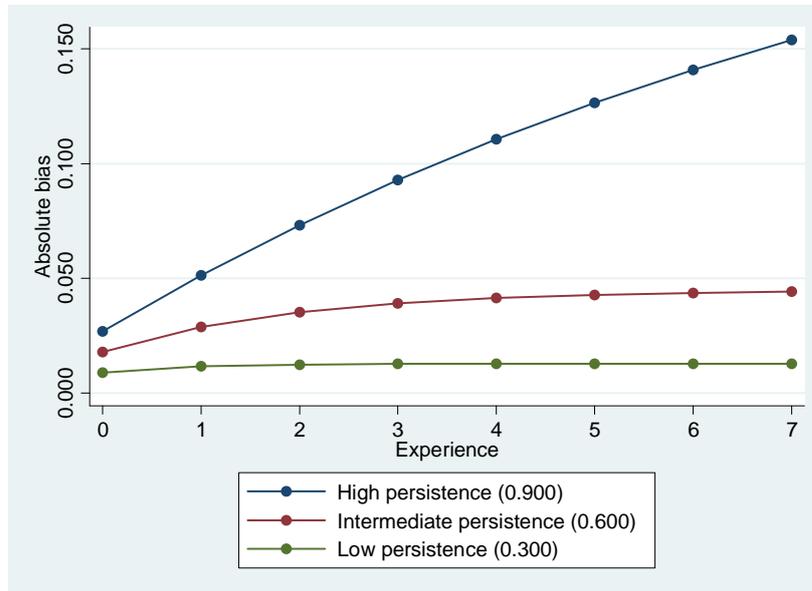
where $z = 0, \dots, T$. This proves that the standard static model (A3) (or (A2)) is not only a particular case ($\theta = 1$) of the more general dynamic model (A7) but also a particular case ($\theta = 1 \Leftrightarrow \rho = 0$) of the more general static model (A8).

Expression (A8) is useful because it allows to see that, in general ($\theta < 1 \Leftrightarrow \rho > 0$), the standard static model (A3) provides a return to schooling (λ) which implicitly assumes $T = \infty$. Indeed, only in the case where $T = \infty$, the return to schooling in model (A8) can be assumed to be independent of z (and equal to a constant number, namely $\frac{\beta}{1-\rho}$). In general, the return to schooling is a function of z .

Expression (A8) also helps to understand why adding simple interactions between schooling and experience in a static model (in order to obtain a z -dependent schooling return) can be misleading.

To conclude, the simplest way to obtain a consistent estimate of the return to schooling with the usually available panel data is to estimate β and ρ separately using the dynamic model (A7) and the GMM-SYS estimation approach described in Andini (2013b). Then, we can use the expression $\beta(\sum_{j=0}^z \rho^j)$ to calculate the return to schooling at each stage of the working life.

Figure 1. Absolute least squares persistence bias



Notes: All simulations assume that $\beta = 0.030$.

Figure 2. Relative least squares persistence bias

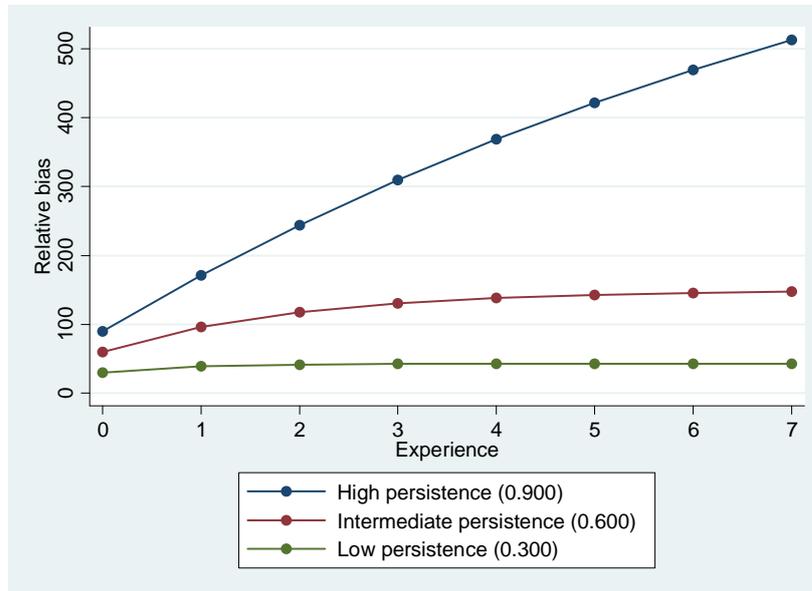
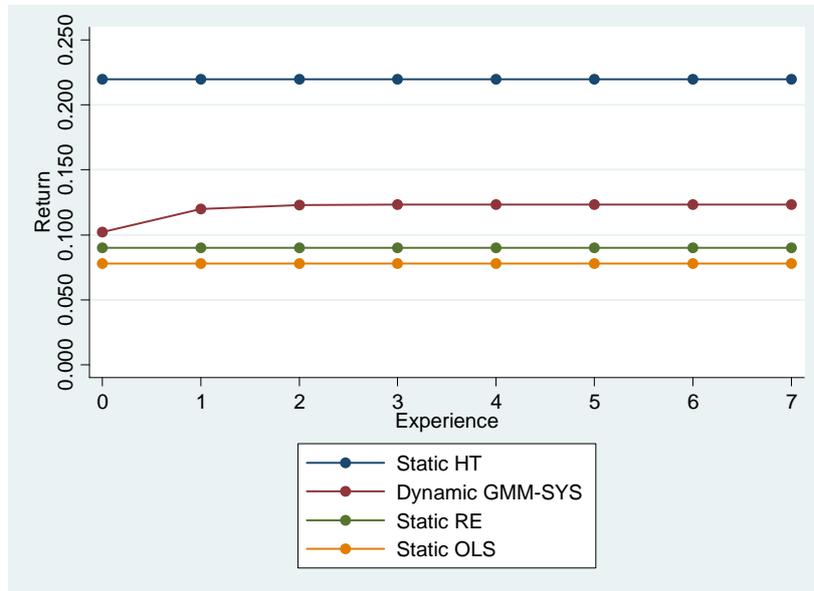


Figure 3. Returns to schooling using alternative approaches



Notes: All estimations are based on the full control set.

Figure 4. The wage-schooling profiles are not parallel across experience groups at the beginning of the working life

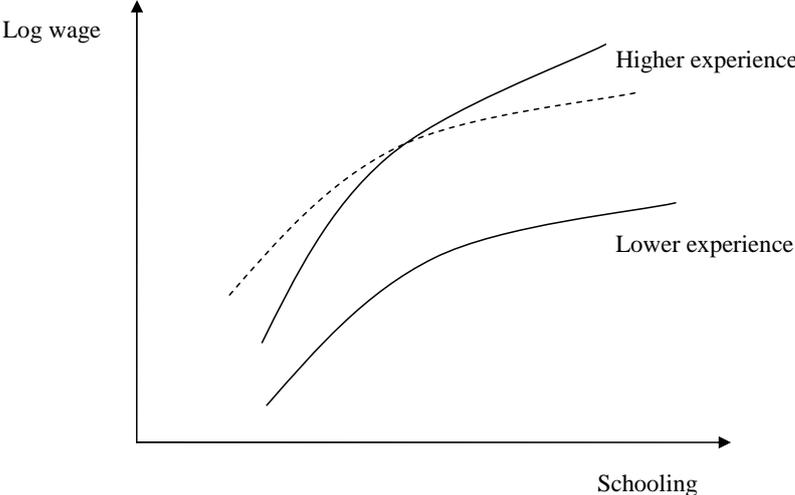


Table 1. Sample descriptive statistics for NLSY data

Variable	Obs	Mean	Std. Dev.	Min	Max
NR	4,360	5,262.059	3,496.150	13	12,548
YEAR	4,360	1983.500	2.291	1980	1987
AGE	4,360	0.032	0.176	0	1
AGE	4,360	24.281	2.772	17	30
BLACK	4,360	0.115	0.319	0	1
BUS	4,360	0.075	0.264	0	1
CON	4,360	0.075	0.263	0	1
ENT	4,360	0.015	0.122	0	1
EXPER	4,360	6.514	2.825	0	18
EXPER2	4,360	50.424	40.781	0	324
FIN	4,360	0.036	0.188	0	1
HISP	4,360	0.155	0.362	0	1
HLTH	4,360	0.016	0.129	0	1
MAN	4,360	0.282	0.450	0	1
MAR	4,360	0.438	0.496	0	1
MIN	4,360	0.015	0.123	0	1
NC	4,360	0.257	0.437	0	1
NE	4,360	0.190	0.392	0	1
OCC1	4,360	0.103	0.305	0	1
OCC2	4,360	0.091	0.288	0	1
OCC3	4,360	0.053	0.224	0	1
OCC4	4,360	0.111	0.314	0	1
OCC5	4,360	0.214	0.410	0	1
OCC6	4,360	0.202	0.401	0	1
OCC7	4,360	0.091	0.289	0	1
OCC8	4,360	0.014	0.120	0	1
OCC9	4,360	0.116	0.321	0	1
PER	4,360	0.016	0.128	0	1
PRO	4,360	0.076	0.265	0	1
PUB	4,360	0.040	0.196	0	1
RUR	4,360	0.203	0.402	0	1
S	4,360	0.350	0.477	0	1
SCHOOL	4,360	11.766	1.746	3	16
TRA	4,360	0.065	0.247	0	1
TRAD	4,360	0.268	0.443	0	1
UNION	4,360	0.244	0.429	0	1
WAGE	4,360	1.649	0.532	-3.579	4.051

NR	Observations number	OCC1	Professional, technical and kindred	AG	Agricultural
YEAR	Year of observation	OCC2	Managers, officials and proprietors	MIN	Mining
AGE	Age	OCC3	Sales workers	CON	Construction
SCHOOL	Schooling years	OCC4	Clerical and kindred	TRAD	Trade
EXPER	Labour-market experience	OCC5	Craftsmen, foremen and kindred	TRA	Transportation
EXPER2	Experience squared	OCC6	Operatives and kindred	FIN	Finance
UNION	Wage set by collective bargaining	OCC7	Labourers and farmers	BUS	Business and repair services
MAR	Married	OCC8	Farm labourers and foreman	PER	Personal services
BLACK	Black	OCC9	Service workers	ENT	Entertainment
HISP	Hispanic			MAN	Manufacturing
HLTH	Has health disability			PRO	Professional and related services
RUR	Lives in rural area			PUB	Public Administration
NE	Lives in North East				
NC	Lives in Northern Central				
S	Lives in South				
WAGE	Log of gross hourly wage				

Table 2. Static vs. dynamic least squares estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS Model (1)	OLS Model (2)	OLS Model (2) Ext 1	OLS Model (2) Ext 2	OLS Model (2) Ext 3	OLS Model (2) Ext 4	OLS Model (2) Full
Control set							
SCHOOL	0.034*** (0.004)	0.076*** (0.004)	0.102*** (0.004)	0.099*** (0.004)	0.093*** (0.004)	0.090*** (0.004)	0.078*** (0.004)
L.WAGE	0.599*** (0.026)						
Observations	3,815	4,360	4,360	4,360	4,360	4,360	4,360
R-squared	0.429	0.064	0.148	0.187	0.204	0.264	0.278
Experience controls			EXPER EXPER2	EXPER EXPER2	EXPER EXPER2	EXPER EXPER2	EXPER EXPER2
Individual controls other than experience				UNION PUB MAR BLACK HISP HLTH	UNION PUB MAR BLACK HISP HLTH	UNION PUB MAR BLACK HISP HLTH	UNION PUB MAR BLACK HISP HLTH
Geographical controls					S NC NE RUR	S NC NE RUR	S NC NE RUR
Industry controls						MIN CON TRAD TRA FIN BUS PER ENT MAN PRO	MIN CON TRAD TRA FIN BUS PER ENT MAN PRO
Occupation controls							OCC1 OCC2 OCC3 OCC4 OCC5 OCC6 OCC7 OCC8

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.10

Table 3. Additional static vs. dynamic least squares estimates

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
	Model (1)	Model (2)	Model (2)	Model (2)
Control set		Full + YE	Full – 80	Full + SZ
SCHOOL	0.034*** (0.004)	0.073*** (0.005)	0.078*** (0.005)	0.100*** (0.011)
L.WAGE	0.599*** (0.026)			
Observations	3,815	4,360	3,815	4,360
R-squared	0.429	0.280	0.270	0.279

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 4. Selected correlations

	L.WAGE	EXPER	EXPER2	YEAR
L.WAGE	1.000			
EXPER	0.149	1.000		
EXPER2	0.109	0.965	1.000	
YEAR	0.239	0.810	0.732	1.000

Table 5. Least squares vs. GMM-SYS estimates

	(1)	(2)	(3)
	OLS Model (1)	OLS Model (1) Full	GMM-SYS Model (1) Full
Control set			
SCHOOL	0.034*** (0.004)	0.037*** (0.004)	0.102*** (0.028)
L.WAGE	0.599*** (0.026)	0.503*** (0.028)	0.174*** (0.031)
Observations	3,815	3,815	3,815
R-squared	0.429	0.469	
IUH accounted	No	No	Yes
Endogeneity accounted	No	No	Yes
Persistence accounted	Yes	Yes	Yes
Number of individuals			545
Number of instruments			171
ABAR1 test (p-value)			0.000
ABAR2 test (p-value)			0.307
Hansen test for all instruments (p-value)			0.246
Difference-in-Hansen test for level equation (p-value)			0.178

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 6. Hausman-Taylor vs. GMM-SYS estimates

	(1)	(2)	(3)	(4)
	RE Model (2) Full	Mundlak Model (2) Full	HT Model (2) Full	GMM-SYS Model (1) Full
Control set				
SCHOOL	0.090*** (0.008)	0.061*** (0.011)	0.220 (0.172)	0.102*** (0.028)
L.WAGE				0.174*** (0.031)
Observations	4,360	4,360	4,360	3,815
IUH accounted	Yes	Yes	Yes	Yes
Endogeneity accounted	No	Partly	Yes	Yes
Persistence accounted	No	No	No	Yes
Number of individuals	545	545	545	545
Number of instruments				171
ABAR1 test (p-value)				0.000
ABAR2 test (p-value)				0.307
Hansen test for all instruments (p-value)				0.246
Difference-in-Hansen test for level equation (p-value)				0.178

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10