Persistence bias and the wage-schooling model

Paper presented at the NOVA School of Business and Economics (Lisbon) on November, 22nd 2013

Corrado Andini

Universidade da Madeira, Campus da Penteada, 9000-390 Funchal, Portugal
Centro de Estudos de Economia Aplicada do Atlântico (CEEAplA), 9501-801 Ponta Delgada, Portugal
Institute for the Study of Labor (IZA), D-53072 Bonn, Germany

ABSTRACT

A well-established empirical literature suggests that individual wages are persistent. Yet, the standard human-capital wage model does not typically account for this stylized fact. This paper investigates the consequences of disregarding earnings persistence when estimating a standard wage-schooling model. In particular, the problems related to the estimation of the schooling coefficient are discussed. Overall, the findings suggest that the standard static-model estimation of the schooling coefficient is subject to persistence bias.

JEL Classification: C23, I21, J31
Keywords: schooling, wages, dynamic panel-data models

Corresponding author:

Prof. Corrado Andini
Universidade da Madeira
Campus da Penteada
9000-390 Funchal
Portugal
E-mail: andini@uma.pt
Tel.: +351 291 70 50 53
Fax: +351 291 70 50 49

Acknowledgments

An earlier version of this manuscript has been published as an IZA discussion paper in January 2013. For valuable comments and suggestions, the author would like to thank Álvaro Novo, Monica Andini, Fabrizio Mazzonna, Massimo Filippini, Vincenzo Galasso and the participants at presentations held in Rome (LUISS, Sep. 2013) and Lugano (USI, 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 that the least squares estimation of the schooling coefficient is subject to two types of bias, 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 (reduce) the Griliches’s biases have been based on three main empirical approaches: extensions of the control set (to proxy unobserved error components and thus reduce the ‘importance’ of the error term), instrumental-variable estimation (to control for endogeneity), and the use of better data (such as 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 a static wage-schooling model 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. Let us start with the least squares case.

To begin with, this paper investigates the consequences of a new (some may say old) type of bias affecting the least squares estimation of the schooling coefficient in a simple wage-schooling model. 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 important source of distortion, the bias arising from the least squares estimation of the schooling coefficient in a static wage-schooling model which disregards earnings persistence. We will refer to it as the ‘least squares persistence bias’.

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. At opposite, if earnings persistence were important, then disregarding such persistence 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 focused on the macroeconomic evidence. The latter has instead 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. In contrast, it dates several decades back. For instance, some of the first articles taking the dynamic aspects of individual earnings models into account have been authored in the 1970s and the 1980s by Lillard and Willis (1978), MaCurdy (1982) and Abowd and Card (1989), among others. More recently, individual-level dynamic wage models taking the persistent nature of earnings into account have been proposed and estimated by Guiso et al. (2005), Cardoso and Portela (2009), and Hospido (2012), to cite a few.

However, despite the existing empirical evidence on the persistence of individual wages, the incorporation of the persistent nature of individual earnings in human-capital or 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 couple of exceptions.

In particular, the importance of accounting for earnings persistence in wage-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. In addition, Andini (2013a; 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 estimation approach suggested by Andini (2013a; 2013b).

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 dynamic nature of the wage-schooling link in the least squares estimation of the schooling coefficient. In addition, this paper goes beyond specific least-squares case by discussing the problems of other static-model estimators: those accounting for endogeneity and 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 the use of the Hausman-Taylor estimator, accounting for both individual unobserved heterogeneity and endogeneity, does not solve the persistence-bias problem.

Specifically, this paper provides 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 potential labor-market experience (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, 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, the third and the fifth of the above results are sample-specific, first and the fourth hold under very general conditions.

In short, the standard cures for the Griliches’s biases (based on extensions of the control set, treatments of endogeneity and models with individual unobserved heterogeneity) are unable to solve the persistence-bias problem related to the estimation of static wage-schooling models. Therefore, an enormous number of schooling coefficient estimates, based on static models, is potentially subject to the persistence-bias critique. Overall, the findings support the dynamic approach to the estimation of wage-schooling models recently suggested by Andini (2013a; 2013b).

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 of the persistence bias of the standard instrumental-variable estimator for the schooling coefficient. Section 6 highlights that disregarding earnings persistence is still problematic even if individual unobserved heterogeneity and endogeneity are accounted for. In particular, the case of the Hausman-Taylor estimator is discussed. Section 7 concludes.

2. Persistence bias in static least squares models

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:

\[
\begin{align*}
 w_{i,s+z+1} &= \alpha + \beta s_i + \rho w_{i,s+z} + u_{i,s+z+1} \quad \text{for } \forall i, s + z \quad \text{with } s \geq 1 \quad z \geq 0
\end{align*}
\]

where \(w\) is logarithm of gross hourly wage, \(s\) is schooling years, \(z\) is years of potential labor-market experience, and \(u\) is an error term. Hence the ‘true’ model is dynamic in
the sense that past wages help to predict current wages. A more general version of
model (1) is described in Appendix A. However, to make the point of this section, we
keep the presentation as simple as possible.

In addition, let us assume that:

(H1) \( \text{COV}(s_i, u_{i,s+z+1}) = 0 \) \( \forall i, s + z \)

(H2) \( \text{COV}(w_{i,s+z}, u_{i,s+z+1}) = 0 \) \( \forall i, s + z \)

(H3) \( \text{COV}(u_{i,s+z}, u_{i,s+z+1}) = 0 \) \( \forall i, s + z \)

(H4) \( \text{COV}(u_{i,s+z}, u_{j,s+z}) = 0 \) \( \forall i \neq j, s + z \)

(H5) \( \text{E}(u_{i,s+z+1}) = 0 \) \( \forall i, s + z \)

(H6) \( \text{VAR}(u_{i,s+z+1}) = \theta^2 \) \( \forall i, s + z \)

(H7) \( \text{VAR}(s_i) = \sigma^2 \) \( \forall i \)

(H8) \( \text{COV}(s_i, \rho w_{i,s-1} + u_{i,s}) = 0 \) \( \forall i, s \)

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 for
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 wage\(^4\) 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, if the ‘true’ model is (1) but earnings persistence is disregarded and the following static ‘false’ model is estimated:

\[
(2) \quad w_{i,s+z+1} = \alpha + \beta s_i + e_{i,s+z+1} \quad \text{where} \quad e_{i,s+z+1} = \rho w_{i,s+z} + u_{i,s+z+1}
\]

then, it is easy to show that:

\[
(3) \quad \lim_{s \to \infty} \beta_{OLS} = \beta + \rho \frac{\text{COV}(s_i, w_{i,s+z})}{\text{VAR}(s_i)}
\]

Knowing that \(\text{VAR}(s_i) = \sigma^2\), it is possible to focus on \(\text{COV}(s_i, w_{i,s+z})\). In particular, it can be shown that:

\[
(4) \quad \text{COV}(s_i, w_{i,s+z}) = \text{COV}(s_i, \alpha + \beta s_i + \rho w_{i,s+z-1} + u_{i,s+z}) = \\
= \beta \sigma^2 + \rho \text{COV}(s_i, w_{i,s+z-1}) = \beta \sigma^2 + \rho \text{COV}(s_i, \alpha + \beta s_i + \rho w_{i,s+z-2} + u_{i,s+z-1}) = \\
= \beta \sigma^2 + \rho \left[ \beta \sigma^2 + \rho \text{COV}(s_i, w_{i,s+z-2}) \right] = \beta \sigma^2 + \rho \beta \sigma^2 + \rho^2 \text{COV}(s_i, w_{i,s+z-2}) = \\
= \beta \sigma^2 (1 + \rho + \rho^2 + \ldots + \rho^{z-1}) + \rho^2 \text{COV}(s_i, w_{i,s})
\]

Since \(\text{COV}(s_i, w_{i,s}) = \text{COV}(s_i, \alpha + \beta s_i + \rho w_{i,s-1} + u_{i,s}) = \beta \sigma^2 + \text{COV}(s_i, \rho w_{i,s-1} + u_{i,s})\) and \(\text{COV}(s_i, \rho w_{i,s-1} + u_{i,s}) = 0\) by assumption, then we get:
\[
\begin{align*}
\text{COV}(s_i, w_{i,s+z}) &= \beta \sigma^2 (1 + \rho + \rho^2 + \ldots + \rho^{z-1}) + \rho^z \beta \sigma^2 = \\
&= \beta \sigma^2 (1 + \rho + \rho^2 + \ldots + \rho^z)
\end{align*}
\]

Hence, using (3), it follows that:

\[
\text{p} \lim \beta_{\text{OLS}} = \beta + \rho \beta \sum \rho^z
\]

where \( \rho \beta \sum \rho^z \) 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 \( \beta \) and \( \rho \) are positive, with the bias being increasing in both \( \rho \) and \( z \). Obviously, we can define the percent (or relative) bias as the ratio between the absolute bias and \( \beta \). The latter is given by \( \rho \sum \rho^z \), thus being independent of \( \beta \).

As a matter of example, Figure 1 illustrates how the persistence bias increases with \( z \) assuming several degrees of earnings persistence and \( \beta = 0.030 \). The upper plot depicts the absolute bias of the schooling coefficient estimated using a static model. The lower plot 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 lower 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 persistence bias worrisome in static least squares models?

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 in which the persistence bias should be lower than in a standard dataset including older workers since the average potential 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 literature (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 Appendix B.

The estimation results, obtained using the least squares estimator, are presented in Table 1. Column 1 shows the estimates from model (1), the ‘true’ dynamic one. The coefficient of schooling $\beta$ is estimated at 0.034, with the degree of earnings persistence $\rho$ 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 potential 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, upper plot), 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 cure the persistence bias in static least squares models?

Columns 3 to 7 gradually extend the static model (2) to investigate whether the 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 potential experience and its square. However, the coefficient of schooling does not decrease, thus indicating that potential experience (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, like the evolution of the R-squared coefficient suggests. 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 2 performs some robustness checks by considering issues associated with i) the presence of year 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 themselves 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 Appendix C), 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. Yet, the discrepancy does not vanish.

Finally, column 4 in Table 2 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 is able provides a coefficient of schooling close to the ‘true’ one, estimated using model (1).

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, the coefficient of schooling estimated using a static specification can only be interpreted as the return to schooling under very unrealistic assumptions (individuals that never die; see Appendix A for details). Hence, to recover the return to schooling, we still need an estimate of the degree of earnings persistence.

5. Persistence bias in static instrumental-variable models

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.

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 variable\(^6\), and a ‘well-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:

\[
\text{(H9) } \text{COV}(s_i, c_i) \neq 0 \quad \forall i \\
\text{(H10) } \text{COV}(s_i, v_{i,s+z+1}) = 0 \quad \forall i, s + z \\
\text{(H11) } \text{COV}(c_i, v_{i,s+z+1}) = 0 \quad \forall i, s + z
\]
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), under the implicit assumption that $e_{i,s+z+1} = \rho w_{i,s+z} + u_{i,s+z+1}$ and $u_{i,s+z+1} = c_i + v_{i,s+z+1}$. 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 $s_i$, but the role played by the past wage in model (1) is disregarded. In short, the researcher assumes that $\rho = 0$ while this hypothesis does not hold true. The standard static instrumental-variable practice is to find a time-invariant instrument $g_i$ such that $\text{COV}(g_i, s_i) \neq 0$ (for instance, the schooling years of the father of the individual $i$). In this case, it is easy to show that:
The conclusion is that, even if the researcher is able to find an instrument satisfying \( \text{COV}(g_i, u_{i,s+z+1}) = 0 \), i.e. the standard instrumental-variable assumption, the instrumental-variable estimator will still be inconsistent as \( \text{COV}(g_i, s_i) \neq 0 \) implies \( \text{COV}(g_i, w_{i,s+z}) \neq 0 \). This is trivial because \( w_{i,s+z} \) is correlated with \( s_i \). The last term of the sum in expression (7) is the absolute ‘instrumental-variable persistence bias’.

This instrumental-variable 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 or human-capital literature. In addition, it is also important for the (strictly-speaking) experimental literature since, as stressed by Carneiro et al. (2006, p. 2), the instrumental-variable method “is the most commonly used method of estimating \( \beta \). 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.

6. Persistence bias in static Hausman-Taylor (panel data) models

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 (2013a; 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 How to obtain consistent estimates: the GMM-SYS estimator

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

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

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

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 schooling coefficient and it has the advantage to be easily testable. In particular, the additional orthogonality conditions imposed by the level equation (9) must pass the Difference-in-Hansen test.

A further requirement is that the level-equation instruments should not be weak. This may happen in presence of non-stationary variables. The latter is also an easily testable assumption. A test can be based on the estimation of an AR1 process (with constant term) for the variable in levels, again using the GMM-SYS estimator. A preliminary 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, it is likely that the true autoregressive coefficient of the log-wage process is well below the critical value of 1.000. Of course, if one or more variables are found to be non-stationary, they should be excluded from the set of level-equation instruments.
Using the full control set, the GMM-SYS estimator provides an estimate of the degree of earnings persistence $\rho$ equal to 0.174 and an estimate of the schooling coefficient $\beta$ equal to 0.102, both significant at 1% level.

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 3 confirms the above view. Column 1 reports the least squares estimates of model (1) with no controls. Column 2 adds all the controls considered in column 7 of Table 1, i.e. the full control set. The finding is that there is no big difference in the estimates of both $\beta$ and $\rho$ between column 1 and column 2. However, once individual unobserved heterogeneity and endogeneity are taken into account using the GMM-SYS estimator, the finding 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. So, the problem with the least squares approach to model (1) is that it does not take into account individual unobserved heterogeneity and endogeneity.

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 4 presents some additional evidence comparing the ‘true’ estimate of the schooling coefficient based on the GMM-SYS estimator, again reported in column 3, 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 4, 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 individual unobserved 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 4, 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. For instance, the mean value of the health status is used as instrument for schooling.

Focusing on the Hausman-Taylor estimation, the conclusion seems to be that again, likewise the classical instrumental-variable case, 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 4. The good news for static-model users is that the GMM-SYS estimate of the schooling coefficient seems to be in line with the random effects estimate (0.090). This can be observed by
comparing column 1 and column 4 in Table 4. In contrast, the schooling coefficient estimated using the Mundlak approach seems to be biased downward.

More interestingly, the static least squares Mincerian model in column 4 of Table 1 seems to provide a very good proxy for the ‘true’ coefficient (0.102), suggesting that, once a quadratic function of experience is accounted for, the least squares estimator may benefit from the possibility that persistence, ability, attenuation and omitted-variable biases compensate each other. Although we are sceptical about the possibility of such a compensation to be systematic, we believe that this finding is something worth mentioning.

6. Conclusions
There are at least three intuitive reasons why wage-schooling models should by 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, representing wage or productivity disturbances, may show some degree of persistence (Guvenen, 2009, among many others, models them as autoregressive of order one). 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 the link between schooling and wages. 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 when estimating a standard wage-schooling model. We have argued that the estimation of the schooling coefficient in a static wage-schooling model is, in general, biased.

Five main results have been presented in this paper. First, the least squares estimator of the schooling coefficient has been shown to be biased upward, with a bias increasing in potential labor-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, we are aware that the second, the third and the fifth of the above findings are specific to our sample. However, we have shown, under very general conditions, that both the least squares estimator and the instrumental-variable estimator produce biased estimates of the schooling coefficient when earnings persistence is disregarded.

Overall, the findings support the dynamic approach to the estimation of wage-schooling models recently proposed by Andini (2013a; 2013b). One very important implication of our findings is that the return to schooling cannot be consistently estimated using a static wage model. If the estimate of the schooling coefficient is biased in the static model, then the estimate of the schooling return is obviously biased too. Indeed, the schooling return should be computed using the dynamic approach described in Andini (2013b). In such dynamic approach, the return to schooling does not generally coincide with the coefficient of schooling. In particular, the schooling return is obtained using estimates of both the degree of earnings persistence and the schooling coefficient (the exact expression is provided by Andini, 2013b). It thus follows that, in order to obtain a consistent estimate of the schooling return, consistent estimates of both the degree of earnings persistence and the schooling coefficient are
needed. Another important implication of a dynamic approach is that, unlike the standard static wage model, the return to schooling is not independent of individual potential labour-market experience (or age). Hence, a dynamic approach allows us to compute the return to schooling at labour-market entry as well as at any specific point in time during the individual working life. The relevance of the above implications for the literature on schooling returns is straightforward.
References


Simulation parameters: $\beta = 0.030$ and $\rho = \begin{cases} 0.900 & \text{ (green)} \\ 0.600 & \text{ (red)} \\ 0.300 & \text{ (blu)} \end{cases}$
<table>
<thead>
<tr>
<th>Control set</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Model (1)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>0.034***</td>
<td>0.076***</td>
<td>0.102***</td>
<td>0.099***</td>
<td>0.093***</td>
<td>0.090***</td>
<td>0.078***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>L.WAGE</td>
<td>0.599***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 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 |

Controls added to model (2) in previous column:
- EXPER
- UNION
- S
- MIN
- OCC1
- EXPER2
- PUB
- NC
- CON
- OCC2
- MAR
- NE
- TRAD
- OCC3
- BLACK
- RUR
- TRA
- OCC4
- HISP
- HLTH
- FIN
- OCC5
- MIN
- BUS
- OCC6
- PER
- OCC7
- OCC8
- MAN
- PRO

Excluded categories: AG, OCC9 and YEAR87
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
### Table 2

<table>
<thead>
<tr>
<th>Control set</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Model (1)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
<td>OLS Model (2)</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>0.034*** (0.004)</td>
<td>0.073*** (0.005)</td>
<td>0.078*** (0.005)</td>
<td>0.100*** (0.011)</td>
</tr>
<tr>
<td>L.WAGE</td>
<td>0.599*** (0.026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3,815</td>
<td>4,360</td>
<td>3,815</td>
<td>4,360</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.429</td>
<td>0.280</td>
<td>0.270</td>
<td>0.279</td>
</tr>
</tbody>
</table>

Excluded categories: AG, OCC9 and YEAR87

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS Model (1)</td>
<td>OLS Model (1)</td>
<td>GMM-SYS Model (1)</td>
</tr>
<tr>
<td>Control set</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHOOL</td>
<td>0.034***</td>
<td>0.037***</td>
<td>0.102***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>L.WAGE</td>
<td>0.599***</td>
<td>0.503***</td>
<td>0.174***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.028)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,815</td>
<td>3,815</td>
<td>3,815</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.429</td>
<td>0.469</td>
<td></td>
</tr>
<tr>
<td>IUH accounted</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Endogeneity accounted</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Persistence accounted</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of individuals</td>
<td></td>
<td></td>
<td>545</td>
</tr>
<tr>
<td>Number of instruments</td>
<td></td>
<td></td>
<td>171</td>
</tr>
<tr>
<td>ABAR1 test (p-value)</td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>ABAR2 test (p-value)</td>
<td></td>
<td></td>
<td>0.307</td>
</tr>
<tr>
<td>Hansen test for all instruments (p-value)</td>
<td></td>
<td></td>
<td>0.246</td>
</tr>
<tr>
<td>Difference-in-Hansen test for level equation (p-value)</td>
<td></td>
<td></td>
<td>0.178</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th>Control set</th>
<th>(1) RE Model (2) Full</th>
<th>(2) Mundlak Model (2) Full</th>
<th>(3) HT Model (2) Full</th>
<th>(4) GMM-SYS Model (1) Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHOOL</td>
<td>0.090*** (0.008)</td>
<td>0.061*** (0.011)</td>
<td>0.220 (0.172)</td>
<td>0.102*** (0.028)</td>
</tr>
<tr>
<td>L.WAGE</td>
<td>0.174*** (0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4,360</td>
<td>4,360</td>
<td>4,360</td>
<td>3,815</td>
</tr>
<tr>
<td>IUH accounted</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Endogeneity accounted</td>
<td>No</td>
<td>Partly</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Persistence accounted</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>545</td>
<td>545</td>
<td>545</td>
<td>545</td>
</tr>
<tr>
<td>Number of instruments</td>
<td></td>
<td></td>
<td></td>
<td>171</td>
</tr>
<tr>
<td>ABAR1 test (p-value)</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
</tr>
<tr>
<td>ABAR2 test (p-value)</td>
<td></td>
<td></td>
<td></td>
<td>0.307</td>
</tr>
<tr>
<td>Hansen test for all instruments (p-value)</td>
<td></td>
<td></td>
<td></td>
<td>0.246</td>
</tr>
<tr>
<td>Difference-in-Hansen test for level equation (p-value)</td>
<td></td>
<td></td>
<td></td>
<td>0.178</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1
Appendix A. A general wage-schooling model

Suppose individual log-productivity \( y_{i,s+z+1} \) is a linear function \( f(.) \) of time-invariant observed schooling years \( s_i \), time-invariant unobserved abilities \( a_i \), which are allowed to be correlated with schooling years, and a set of other time-varying observed factors \( X_{i,s+z+1} \), including potential labour-market experience \( z \). In short, we have:

\[
(A1) \quad y_{i,s+z+1} = \pi a_i + \alpha s_i + \delta X_{i,s+z+1}
\]

The standard human-capital theory suggests that:

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

(Standard model, implicit version)

or alternatively:

\[
(A3) \quad w_{i,s+z+1} = \pi a_i + \alpha s_i + \delta X_{i,s+z+1} + v_{i,s+z+1}
\]

(Standard model, explicit version)

where the residuals are 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 \( \theta = 1 \).

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

(Adjustment model)

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

(Wage-bargaining model)

\[
(A6) \quad w_{i,s+z+1} = y_{i,s+z+1} + (1-\theta)v_{i,s+z} + v_{i,s+z+1}
\]

(Autocorrelated-disturbances model)

For a discussion about (A4), see Andini (2010; 2013b). For a discussion about (A5), see Andini (2009; 2013a). For a discussion about (A6), see Guvenen (2009), among others.

Further, the above three models can be all written as one single model, by appropriately re-labelling parameters \( \rho = 1 - \theta \), \( \beta = \theta \alpha \), \( \gamma = \theta \delta \), \( c_i = \theta \pi a_i \):

\[
(A7) \quad w_{i,s+z+1} = \rho w_{i,s+z} + \beta s_i + \gamma X_{i,s+z+1} + c_i + v_{i,s+z+1}
\]

(General model, dynamic version)

This is the general wage-schooling model referred in the title of this appendix. Of course, this model can be made even more general by allowing for a dynamic discrete-choice model of schooling decisions, in the spirit of the ‘structural’ literature (see footnote 2). The coefficient of schooling in the static model (\( \alpha \)) only coincides with
that of the dynamic model ($\beta = 0\alpha$) in a very special case ($\theta = 1$). In general ($\theta < 1$), it is higher ($\alpha > \beta$).

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

\[
(A8) \quad w_{i,s+z} = \rho^{z+1}w_{i,s-1} + \beta \left( \sum_{j=0}^{\infty} \rho^j x_{i,s+z-j} \right) + \gamma \left( \sum_{j=0}^{\infty} \rho^j X_{i,s+z-j} \right) + \sum_{j=0}^{\infty} \rho^j v_{i,s+z-j}
\]

(General model, static version)

where $z = 0, \ldots, T$. Thus, the return to schooling is a function of $z$. By assuming it constant $\left( \frac{\beta}{1-\rho} = \alpha \right)$ over the working life, the standard model is implicitly assuming that individuals have infinite potential labour-market experience (i.e. they never die). 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). In general ($\theta < 1 \Leftrightarrow \rho > 0$), the standard static model (A3) provides a return to schooling $\left( \frac{\beta}{1-\rho} = \alpha \right)$ which is above the ‘true’ one $\left( \beta \left( \sum_{j=0}^{\infty} \rho^j \right) \right)$, unless $z = \infty$.

Expression (A8) also helps to show that, even if we allow for the schooling coefficient to depend on $z$ in the static model (by adding interactions between schooling and experience), there is no way to obtain a consistent estimate of the return to schooling at each $z$ level, with the usually available panel data and using static-model estimation techniques. Indeed, the observation of the $X$ variables is generally limited in time, i.e. we do not typically have data on the whole history of the individual characteristics. In addition, the standard practice is even less appropriate because the lagged values of the $X$ variables are usually disregarded and the errors are assumed to be i.i.d. (see model (A3)).

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

The data are taken from the National Longitudinal Survey of Youth. The dataset contains observations on 545 males for the period of 1980-1987. The statistics of the variables and their meaning are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>4360</td>
<td>5262.059</td>
<td>3496.150</td>
<td>13</td>
<td>12548</td>
</tr>
<tr>
<td>AG</td>
<td>4360</td>
<td>0.032</td>
<td>0.176</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BLACK</td>
<td>4360</td>
<td>0.115</td>
<td>0.319</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BUS</td>
<td>4360</td>
<td>0.075</td>
<td>0.264</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CON</td>
<td>4360</td>
<td>0.075</td>
<td>0.263</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ENT</td>
<td>4360</td>
<td>0.015</td>
<td>0.122</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EXPER</td>
<td>4360</td>
<td>6.514</td>
<td>2.825</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>EXPER2</td>
<td>4360</td>
<td>50.424</td>
<td>40.781</td>
<td>0</td>
<td>324</td>
</tr>
<tr>
<td>FIN</td>
<td>4360</td>
<td>0.036</td>
<td>0.188</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HISP</td>
<td>4360</td>
<td>0.155</td>
<td>0.362</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HLTH</td>
<td>4360</td>
<td>0.016</td>
<td>0.129</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MAN</td>
<td>4360</td>
<td>0.282</td>
<td>0.450</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MAR</td>
<td>4360</td>
<td>0.438</td>
<td>0.496</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MIN</td>
<td>4360</td>
<td>0.015</td>
<td>0.123</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NC</td>
<td>4360</td>
<td>0.257</td>
<td>0.437</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NE</td>
<td>4360</td>
<td>0.190</td>
<td>0.392</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC1</td>
<td>4360</td>
<td>0.103</td>
<td>0.305</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC2</td>
<td>4360</td>
<td>0.091</td>
<td>0.288</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC3</td>
<td>4360</td>
<td>0.053</td>
<td>0.224</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC4</td>
<td>4360</td>
<td>0.111</td>
<td>0.314</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC5</td>
<td>4360</td>
<td>0.214</td>
<td>0.410</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC6</td>
<td>4360</td>
<td>0.202</td>
<td>0.401</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC7</td>
<td>4360</td>
<td>0.091</td>
<td>0.289</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC8</td>
<td>4360</td>
<td>0.014</td>
<td>0.120</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>OCC9</td>
<td>4360</td>
<td>0.116</td>
<td>0.321</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PER</td>
<td>4360</td>
<td>0.016</td>
<td>0.128</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PRO</td>
<td>4360</td>
<td>0.076</td>
<td>0.265</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PUB</td>
<td>4360</td>
<td>0.040</td>
<td>0.196</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>RUR</td>
<td>4360</td>
<td>0.203</td>
<td>0.402</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>4360</td>
<td>0.350</td>
<td>0.477</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>4360</td>
<td>11.766</td>
<td>1.746</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>TRA</td>
<td>4360</td>
<td>0.065</td>
<td>0.247</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TRAD</td>
<td>4360</td>
<td>0.268</td>
<td>0.443</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>UNION</td>
<td>4360</td>
<td>0.244</td>
<td>0.429</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WAGE</td>
<td>4360</td>
<td>1.649</td>
<td>0.532</td>
<td>-3.579</td>
<td>4.051</td>
</tr>
</tbody>
</table>

Observational dummies:
- OCC1: Professional, technical and kindred
- OCC2: Managers, officials and proprietors
- OCC3: Sales workers
- OCC4: Clerical and kindred
- OCC5: Craftsmen, foremen and kindred
- OCC6: Operatives and kindred
- OCC7: Laborers and farmers
- OCC8: Farm laborers and foreman
- OCC9: Service workers
- AG: Agricultural
- MIN: Mining
- CON: Construction
- TRAD: Trade
- TRA: Transportation
- FIN: Finance
- BUS: Business and repair services
- PER: Personal services
- ENT: Entertainment
- MAN: Manufacturing
- PRO: Professional and related services
- PUB: Public Administration

Industry dummies:
- NR: Observations number
- YEAR: Year of observation
- SCHOOL: Schooling years
- EXPER: Potential labor-market experience
- EXPER2: Experience squared
- UNION: Wage set by collective bargaining
- MAR: Married
- BLACK: Black
- HISP: Hispanic
- HLTH: Has health disability
- RUR: Lives in rural area
- NE: Lives in North East
- NC: Lives in Northern Central
- S: Lives in South
- WAGE: Log of gross hourly wage
Appendix C. Selected correlations

The correlation matrix for selected variables in the dataset is the following:

<table>
<thead>
<tr>
<th></th>
<th>L.WAGE</th>
<th>EXPER</th>
<th>EXPER2</th>
<th>YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.WAGE</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPER</td>
<td>0.149</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPER2</td>
<td>0.109</td>
<td>0.965</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>YEAR</td>
<td>0.239</td>
<td>0.810</td>
<td>0.732</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Endnotes

1 Some 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.

2 As 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 which is 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. Indeed, as shown in Appendix A, the standard model estimated in the instrumental-variable literature can be seen as a particular case of a more general wage-schooling model, either dynamic or static. For sake of clarification, our paper 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 past wages do not play any explicit role, we do not model schooling decisions (likewise the instrumental-variable approach) but we see an explicit role for past wages (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.

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

4 The idea of a reservation wage is compatible with the presence of self-selection into the labor 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).

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

6 If 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 \( c_1 \) can be assumed to capture this type of individual unobserved heterogeneity.

7 Another 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 Vytlacil, 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.

8 One limitation of the approach proposed by Andini (2013a; 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).