Working Paper 7/01

THE ERROR STRUCTURE OF EARNINGS: AN ANALYSIS ON ITALIAN LONGITUDINAL DATA

Margherita Borella[♣]

♣ CeRP and University College London
Via Real Collegio, 30
10024 Moncalieri (Turin)
Tel.: +39.011.6476525
Fax: +39.011.6403680
E.mail: borella@cerp.unito.it

ABSTRACT

The aim of this paper is to characterise the time series properties of earnings in Italy, using the panel data set drawn from the Bank of Italy Survey of Households' Income and Wealth (SHIW). The Bank of Italy Survey is drawn every two years: this feature raises identification problems as the first-order autocovariance is not observed. However, it is possible to use the panel dimension of the data set in order to discriminate between several specifications that imply different covariance patterns. In order to exploit the differences that may arise due to heterogeneous education attainments, estimates are performed by education group.
Results show that the AR(1) plus individual effect model provides the best characterisation of the unobserved component of the earnings process. The estimated autoregressive parameter however is well below unity, indicating stationarity.

I am very grateful to Orazio Attanasio and Elsa Fornero for their help and comments, and to Fabio Bagliano, Alessandro Sembenelli and participants to CeRP seminars for useful comments. I am indebted to Agar Brugiavini who kindly provided me with the algorithm used to build gross earnings.

1. Introduction

The availability of longitudinal surveys has allowed researchers to model the individuals' covariance pattern of earnings over time. Several authors using US panel data have performed this kind of study[1]. In particular, MaCurdy (1982) develops a set of statistical procedures in order to choose among different specifications of the error structure. In his application to the Michigan Panel of Income Dynamics, his preferred specification is given by an MA(2) model applied to the change in (the logarithm of) earnings, which implies an ARMA(1,2) model with a unit root for the same variable expressed in levels.

Using the Panel Study of Income Dynamics (PSID), Moffitt and Gottschalk (1995) model the unobserved component of (the logarithm of) earnings as the sum of a transitory component and a permanent component; in their preferred specification the permanent component is modelled as a random walk process.

In this paper, the same line of research is followed in order to characterise the time series properties of earnings in Italy, using the panel data set drawn from the Bank of Italy Survey of Households' Income and Wealth (SHIW). The Bank of Italy Survey is drawn every two years: this feature raises identification problems as the first-order autocovariance is not observed. It is therefore not possible to distinguish among stationary models that imply one-lag covariances in the structure, as would be the case of an MA(1) component. However, it is possible to use the panel dimension of the data set in order to discriminate between several specifications that imply different covariance patterns. In particular, it is possible to characterize both the standard permanent-transitory model and models that contain AR(1) components. In addition, in order to exploit the differences that may arise due to heterogeneous education attainments, estimates are performed by education group.

Results show that the AR(1) plus individual effect model provides the best characterisation of the unobserved component of the earnings process. The estimated autoregressive parameter however is well below unity, indicating stationarity.

Section 2 develops the theoretical models that will be tested in the empirical analysis, section 3 gives a brief description of the data set, and section 4 presents the results. Section 5 concludes.

---

[1] Among others, Lillard and Willis (1978), Abowd and Card (1989), and Gottschalk and Moffitt (1994) use the PSID data set in order to characterise the earnings process.

2. Models for the Earning Process

The empirical formulation for the earning process typically used in the literature[2] is:

$$y_{it}^{a} = X_{it}^{a}{}'\beta + u_{it}^{a}$$

y_{it}^{a} is the natural logarithm of real earnings of the *i*-th individual at time *t*, where the index *a* (age) has been added to stress the fact that the variables in the model may as well depend on the position of the individual over the life-cycle. X_{it}^{a} is a (k×1) vector of observable variables, β is a (k×1) vector of unknown parameters, and u_{it}^{a} is an error term which represents unobserved characteristics determining earnings. The variables included in *X* are a polynomial in age, which captures the life-cycle profile of earnings, measures of education and other information available about the labour supply behaviour of the individuals in the sample. In addition, time dummies for each period are included in order to capture the common period effects. Consequently, the disturbances u_{it}^{a} are assumed to be independently distributed across individuals but not over time. Modelling their covariance structure is the main concern of this study. Several specifications have been proposed and tested in the literature: here the attention is concentrated on those specifications that can be identified using the Bank of Italy panel data set, which collects data every two years.

Permanent-Transitory Model

The simplest model for the earnings structure that has been studied in the literature is the permanent-transitory model, where the unobserved component of earnings for an individual *i* of age *a* is decomposed into a permanent component which is time invariant (μ_i) and a transitory idiosyncratic shock (ε_{it}^{a}).

$$u_{it}^{a} = \mu_i + \varepsilon_{it}^{a} \qquad (1)$$

where both μ_i and ε_{it}^{a} are i.i.d. with zero mean and variance equal to σ_{μ}^{2} and σ_{ε}^{2}, respectively.

---

[2] Among others, see Lillard and Willis (1978), MaCurdy (1982), Abowd and Card (1989) and Moffitt and Gottschalk (1995).

Estimation of this model is feasible if one observes the variance of earnings and its covariance. The theoretical moments are:

$$\text{var}(u_{it}^{a}) = \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2} \qquad \text{and}$$

$$\text{cov}(u_{it}^{a},\, u_{i,t-s}^{a-s}) = \sigma_{\mu}^{2} \qquad s=1,\, 2,\ldots.$$

The major implications of this model are that the variance and the covariances of the unobserved component of earnings are constant over time. In addition, the theoretical covariances are identical at different lags. From those conditions it is clear that the model can be identified observing, in addition to the variance, the covariances at lag 2, 4, and so on. Therefore the parameters in the model can be identified using the Bank of Italy panel data set.

<u>More realistic models</u>

A model that has proved to be a good characterization of the earning process in the US is a model where the transitory component exhibits some autocorrelation. Assuming the transitory component follows an AR(1) process, the unobserved component of earnings can be written as:

$$\begin{aligned} u_{it}^{a} &= \mu_i + z_{it}^{a} \\ z_{it}^{a} &= \alpha\, z_{i,t-1}^{a-1} + \omega_{it}^{a} \end{aligned} \qquad (2)$$

where ω_{it}^{a} is an i.i.d. stochastic process with zero mean and variance σ_{ω}^{2} and μ_i is defined as before. This structure can be estimated if one observes the variance of earnings for a given individual and its covariance at different ages and points in time. The variances of this process can be summarised by the following recursions:

$$Var(u_{it}^{a}) = \sigma_{\mu}^{2} + Var(z_{it}^{a})$$

where:

$$Var(z_{it}^{\underline{a}}) = \sigma_{\underline{a}}^{2}$$

and:

$$Var(z_{it}^{a}) = \alpha^2 \, Var(z_{it-1}^{a-1}) + \sigma_{\omega}^{2} \qquad \text{with } a > \underline{a}$$

Similarly, the covariances are defined as:

$$Cov\left(u_{it}^{a}, u_{it-s}^{a-s}\right) = \sigma_{\mu}^{2} + \alpha^{s} \, Var(z_{it-s}^{a-s}) \qquad s=1, 2, \ldots$$

where the formulas reflect the fact that the AR(1) component arises from a finite process starting at age $\underline{a}$, the age at which individuals enter the labour market. Estimation of a finite process allows to overcome the problems associated with unit roots, as the recursion formulas are well defined even if the autoregressive parameter is equal to (or greater than) one.

Contrary to standard time series analysis, the initial values of the autoregressive component should not be treated as known constants in models for longitudinal data where the time dimension is typically quite small[3]. Here the autoregressive process is assumed to start at age $\underline{a}$, and the variance of the zero mean initial distribution of the process z_{it}^{a} ($\sigma_{\underline{a}}^{2}$) is estimated.

A generalization of the autoregressive model just discussed is a model in which the transitory component ω_{it}^{a} is not i.i.d. but displays some autocorrelation. To take a concrete example, consider the case in which ω_{it}^{a} is an MA(1) process:

$$\omega_{it}^{a} = \xi_{it}^{a} + \theta \xi_{i,t-1}^{a-1} \qquad (3)$$

The theoretical moments implied by this structure are shown in the appendix. It should be noticed that the autocovariance function of an ARMA(1,1) model depends on the MA parameter at lags greater than one. However, failure to observe the first order autocovariance may render the empirical identification of such a parameter more difficult to achieve[4].

Estimation is carried out using the minimum distance method, which compares the sample

[3] See Anderson and Hsiao (1982) and MaCurdy (1982).

[4] In order to ease identification, the initial values of the process in this case have been set equal to zero.

moments to the theoretical ones (Chamberlain, 1984). Denoting the ($m \times 1$) vector of sample moments as π and the vector of theoretical moments as $\pi(\alpha)$, which depends on ($n \times 1$) unknown parameters (with $n<m$), the minimum distance method minimizes the function:

$$\min_{\alpha} (\pi - \pi(\alpha))' V (\pi - \pi(\alpha))$$

where V is a weighting matrix. When V is taken to be the inverse of the matrix of fourth moments the estimator is the well-known optimal minimum distance (OMD). However, Altonji and Segal (1996) warn about the bias that arises when estimating covariance structures of this type, and suggest the use of the equally weighted minimum distance (EWMD) which replaces V with the identity matrix. The latter strategy will be used in the estimation.

3. The Data

In order to model the earnings structure and its time series properties, the panel sample from the Bank of Italy Survey has been used. This is the most comprehensive survey of individual data in Italy and it contains detailed information on household members' demographic characteristics and labour supply variables. The Survey has been run since 1977, but it has a panel dimension only since 1989. Data are available until 1998 so that there are 5 consecutive waves of the sample that can be used in estimation[5].

Each wave about 8,000 families representative of the Italian population are interviewed; approximately 40% of them are interviewed in subsequent waves. However, only 10% of households interviewed in 1989 have been interviewed up to 1995. Therefore, the sample used in the analysis has been built using all individuals who have been interviewed for at least two consecutive waves of the survey. The use of an unbalanced sample in estimation considerably reduces the sample attrition bias present in panel data sets.

The dependent variable used in the analysis is built upon the logarithm of real annual gross earnings of each individual in the sample who reported positive earnings and classified

[5] The available years are: 1989, 1991, 1993, 1995 and 1998. This implies that it is possible to compute the sample covariances of order two, three, four, five and so on.

himself as dependent worker (either in the private or in the public sector)[6]. Annual gross earnings have been deflated using the ISTAT consumer price index, and they are expressed in 1998 prices. The analysis is carried out using only male workers aged between 22 and 60, as for male workers the participation issue is less stringent than for female workers. After applying the selection criteria, the overall sample consists of 5,231 observations, of which 3,329 employed in the private sector.

The variable actually used in the analysis is built as the residuals from regressions of the logarithm of gross earnings on a polynomial in age and cohort dummies, controlling for education. In particular, the sample has been divided into 6 year-of-birth groups, in order to remove cohort effects in the variable of interest[7]. The youngest cohort is formed by individuals born between 1963 and 1967 included, and the eldest by individuals born between 1938 and 1942 included. In the analysis, individuals in the youngest cohort are considered as aged 24 in 1989, 26 in 1991 and so on. The other cohorts are treated similarly. Regressions are then performed by education group using as regressors a polynomial in age and cohort and time dummies, both for private and for public employees[8].

Estimates of the different specifications for the unobserved component of earnings are computed splitting the residuals into four groups, arising from two education groups for each sector, public and private. The two education groups are: high school dropouts (2864 observations, of which 2106 employed in the private sector) and high school and college graduates (2267 observations, of which 1223 employed in the private sector). College graduates on their own would form a sample of 222 and 425 observations in the private and in the public sector respectively, which has been considered too small to be treated separately in the analysis.

Figures 1 and 2 plot the sample variance and the second- and fourth-order covariance of the (residuals of) gross earnings both for private and for public sector dependent workers against age. Both figures show that variances and covariances do not appear to increase

---

[6] Earnings are gross of income tax but net of Social Security contributions. The variable actually reported in the Survey is "normal annual net earnings". However, as detailed demographic information is available in the data set, gross earnings have been computed for each individual in the sample.

[7] The quantitative importance of the cohort effects in the cross-sectional variance of earnings has been documented for example by Deaton and Paxon (1994) and Storesletten et al. (2000).

[8] It is not possible to separately identify age, cohort and time effects without any further assumption, as they are linear combinations of one another. In the analysis it is therefore assumed

over time, a feature that is captured by stationary models. In addition, covariances of increasing order appear to decrease slowly, indicating some persistence in unobservable earnings.

4. Results

Minimum distance estimation of the models described above has been performed separately for private and public sector workers. In addition, estimates for different education groups are reported.

Tables 1 and 2 report estimates for the permanent-transitory model described by equation (1) respectively for private sector and public sector employees. Each table reports estimates both for the whole sample and for the two education groups: 1) high school dropouts, and 2) high school and college graduates. Similarly, tables 3 and 4 show estimated coefficients for the AR(1) model with fixed effect, and tables 5 and 6 present estimates of the parameters for the ARMA(1,1) model. In addition, for each table a Wald test is reported, built on the null hypothesis that the parameters are not statistically different in the two sub-samples considered[9].

For private sector dependent workers, the parameter estimates of the permanent-transitory model in table 1 imply that the overall variance of the unobserved component for the entire sample is 0.077, with a permanent variance of 0.036. Columns 2 and 3 in table 1 show the estimated coefficients for the two education groups considered: high school dropouts and high school and college graduates. The Wald statistics, however, indicates that the differences in the estimates are not statistically significant.

In table 2 estimates for the public sector dependent workers of permanent-transitory model are reported. The overall variance estimated for the whole sample is 0.058, expectedly lower than the overall variance for private sector employees. In particular, the overall variance for high school dropouts is 0.044, while for high school and college graduates is 0.066. The Wald statistics indicates that the parameter estimates for the two groups are in this case statistically different from each other.

Turning to the AR(1) estimates, table 3 shows that in the private sector the autoregressive

---

that the time effects are orthogonal to a time trend and add up to zero.

[9] See Appendix for more details.

parameter is statistically different from zero. The value of α for the whole sample is equal to 0.54, a value that indicates stationarity of the estimated process. The fixed effect variance is also estimated to be different from zero. Differences in the estimates of the two groups are not statistically significant.

The AR(1) model for the public sector employees is presented in table 4. The autoregressive parameter is precisely estimated and is higher then the parameter estimated for private sector dependent workers. The variance of the permanent component is not statistically different from zero. The Wald statistics suggests that differences in the parameters of the two groups are statistically significant.

Estimates of the ARMA(1,1) model plus a fixed effect are shown in tables 5 and 6 for private and public sector workers respectively. The moving average parameter is not statistically different from zero, while the other parameter estimates are close to those obtained for the AR(1) representation. In addition, the residual sums of squares are very close for the two models.

This evidence suggests that, given the data set used, the best characterization for the unobserved component of earnings in Italy seems to be represented by the sum of a stationary AR(1) model and a fixed effect.

5. Conclusions

In this study the panel drawn from the Bank of Italy Survey of Households' Income and Wealth has been used in order to characterise the covariance structure of the unobserved component of earnings.

Various models have been estimated for different sectors (private and public) and for different education groups, in order to exploit the differences that may arise due to heterogeneous education attainments. The specification that better captures the features of the data is a model given by the sum of an AR(1) component and an individual fixed effect. The autoregressive coefficient has been estimated to be around 0.55 in the private sector and 0.8 in the public sector. In the latter group, parameter differences among education groups are found statistically significant, while in the former differences in the estimated parameters for the two education groups do not appear to be statistically significant.

References

Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, 57, 411-445.

Altonji, J. G. and L. M. Segal (1996), "Small Sample Bias in GMM Estimation of Covariance Structures", *Journal of Business and Economic Statistics*, 14(3), 353-366.

Anderson, T. W. and C. Hsiao (1981), "Estimation of Dynamic Models with error components", *Journal of the American Statistical Association*, 76, 598-606.

Anderson, T. W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics*, 18, -82

Chamberlain, G. (1984), "Panel Data", in *Handbook of Econometrics*, eds. Z. Griliches and M. Intrilligator. Amsterdam: North Holland.

Deaton, A. and C. Paxon (1994), "Intertemporal choice and inequality", *Journal of Political Economy*, 102, 437-467

Gottschalk, P. and R. A. Moffitt (1994), "The Growth of Earnings Instability in the U.S. Labor Market", *Brooking Papers on Economic Activity*, 2:1994, 217-272.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50(4), 1029-54.

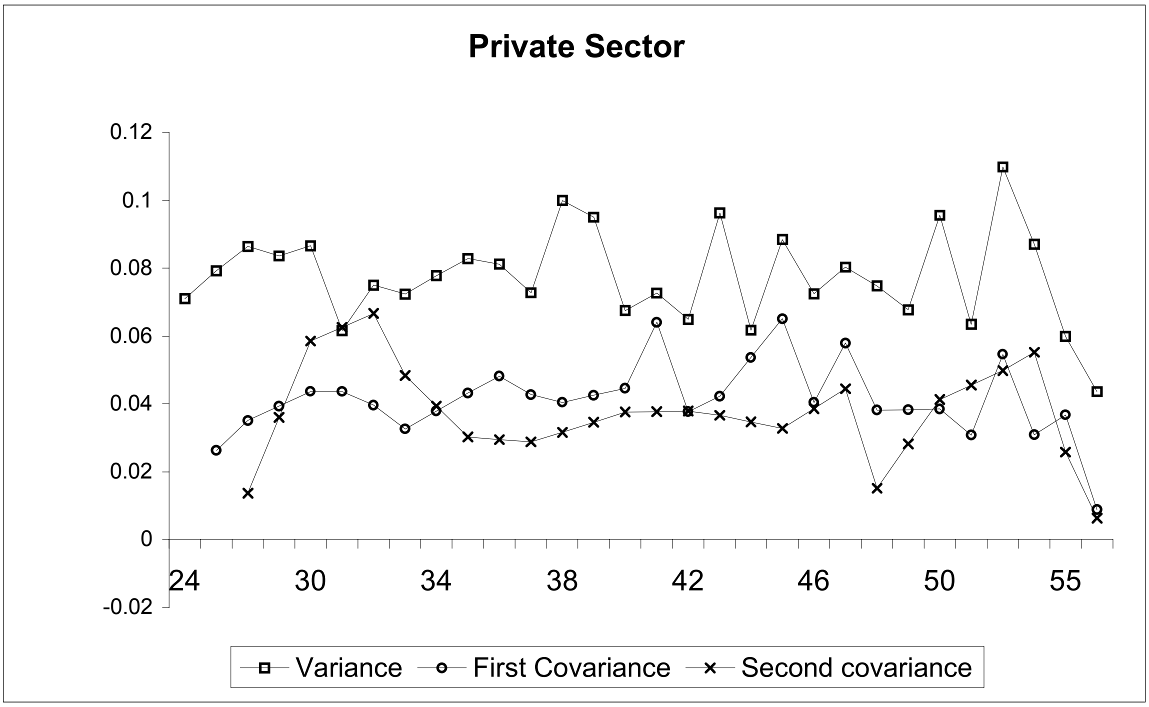
MaCurdy, T. E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics*, 18, 83-114.

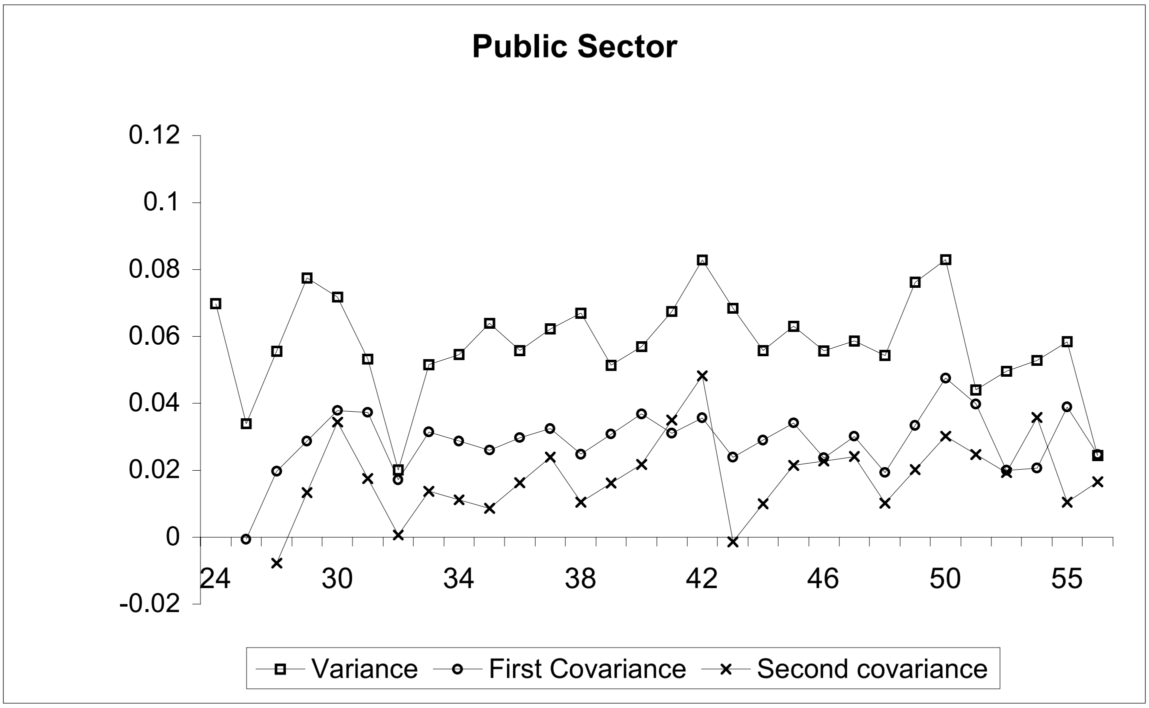
Moffitt, R. A. and P. Gottschalk (1995), "Trends in the Covariance Structure of Earnings in the U.S.: 1969-1987", mimeo.

Lillard, L. A. and R. J. Willis (1978), "Dynamic Aspects of Earning Mobility", *Econometrica*, 46, 985-1012.

Storesletten, K., C. I. Telmer, and A. Yaron (2000), "Consumption and Risk Sharing Over the Life Cycle", mimeo.

Figure 1: Variance and covariances of detrended log earnings.





Permanent-Transitory Model

Table 1: Private sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{μ}^{2}	0.036	0.034	0.039
(t-stat)	(21.22)	(16.57)	(12.46)
σ_{ε}^{2}	0.041	0.041	0.041
(t-stat)	(15.24)	(12.01)	(8.97)
Sum of Squared Residuals	*0.0214*	*0.0240*	*0.0572*
N. of Obs.	*3329*	*2106*	*1223*

Wald statistic: $\chi^2(2)=3.16$

Table 2: Public sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{μ}^{2}	0.020	0.016	0.023
(t-stat)	(10.08)	(9.43)	(8.57)
σ_{ε}^{2}	0.038	0.028	0.043
(t-stat)	(13.00)	(9.57)	(10.52)
Sum of Squared Residuals	*0.0186*	*0.0180*	*0.0321*
N. of Obs.	*1902*	*758*	*1144*

Wald statistic: $\chi^2(2)=36.76$

AR(1) Model

Table 3: Private sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{η}^{2}	0.028	0.027	0.032
(t-stat)	(6.25)	(5.55)	(3.88)
α	0.573	0.545	0.533
(t-stat)	(7.26)	(5.41)	(3.23)
σ_{ω}^{2}	0.033	0.034	0.034
(t-stat)	(11.12)	(8.91)	(6.87)
$\sigma_{\underline{a}}^{2}$	0.037	0.032	0.045
(t-stat)	(2.79)	(1.94)	(1.94)
Sum of Squared Residuals	*0.0199*	*0.0229*	*0.0558*
N. of Obs.	*3329*	*2106*	*1223*

Wald statistic: $\chi^2(4)=3.74$

Table 4: Public sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{η}^{2}	-0.010	0.004	-0.060
(t-stat)	(-0.60)	(0.52)	(-1.07)
α	0.809	0.712	0.908
(t-stat)	(11.65)	(7.70)	(17.40)
σ_{ω}^{2}	0.024	0.021	0.022
(t-stat)	(8.84)	(6.59)	(7.24)
$\sigma_{\underline{a}}^{2}$	0.046	0.018	0.098
(t-stat)	(1.64)	(1.71)	(1.42)
Sum of Squared Residuals	*0.0147*	*0.0147*	*0.0259*
N. of Obs.	*1902*	*758*	*1144*

Wald statistic: $\chi^2(4)=9.89$

ARMA(1,1) Model

Table 5: Private sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{η}^{2}	0.028	0.027	0.033
(t-stat)	(6.68)	(5.15)	(4.16)
α	0.619	0.645	0.473
(t-stat)	(6.30)	(6.92)	(0.93)
θ	-0.200	-0.360	0.097
(t-stat)	(-0.59)	(-0.79)	(0.08)
σ_{ω}^{2}	0.029	0.025	0.036
(t-stat)	(3.53)	(2.26)	(3.50)
Sum of Squared Residuals	*0.0199*	*0.0227*	*0.0559*
N. of Obs.	*3329*	*2106*	*1223*

Wald statistic: $\chi^2(4)=1.98$

Table 6: Public sector dependent workers

	Whole sample	Education: no high school	Education: high school and college
σ_{η}^{2}	0.001	0.003	-0.002
(t-stat)	(0.18)	(0.51)	(-0.15)
α	0.765	0.709	0.795
(t-stat)	(11.64)	(7.31)	(15.31)
θ	-0.204	0.119	-0.247
(t-stat)	(-0.65)	(0.31)	(-0.64)
σ_{ω}^{2}	0.023	0.021	0.024
(t-stat)	(3.65)	(6.77)	(2.92)
Sum of Squared Residuals	*0.0149*	*0.0147*	*0.0271*
N. of Obs.	*1902*	*758*	*1144*

Wald statistic: $\chi^2(4)=2.59$

Appendix

ESTIMATION

Estimation is carried out using the minimum distance method, which compares the sample moments to the theoretical ones (Chamberlain, 1984).

The sample moments are built using the residuals of log earnings as described in section 3 in the text. Denoting the ($m\times1$) vector of sample moments as π and the vector of theoretical moments as $\pi(\alpha)$, which depends on ($n\times1$) unknown parameters (with $n<m$), the minimum distance method minimizes the function:

$$\min_{\alpha} (\pi - \pi(\alpha))' V (\pi - \pi(\alpha))$$

where V is a weighting matrix. Following the findings in the study by Altonji and Segal (1996) on the bias that arises when estimating covariance structures of this type, the identity matrix has been used in estimation, i.e. $V=I$. The equally weighted minimum distance estimator obtained has the following distribution[10]:

$$\sqrt{N}\left(\alpha_{EWMD} - \alpha\right) \xrightarrow{d} N(0, W)$$

The variance-covariance matrix is defined as:

$$W = \left(G'G\right)^{-1} G'VG \left(G'G\right)^{-1}$$

where G is a $(m \times n)$ matrix of first derivatives, and V is the $(m \times m)$ variance-covariance matrix of the moments considered. Each element in V is computed using the residuals for each observation i:

$$v_{m,m'} = \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\pi_{im} - \pi_m(\hat{\alpha}_{EWMD})\right)\left(\pi_{im'} - \pi_{m'}(\hat{\alpha}_{EWMD})\right)\right)$$

As the panel is unbalanced, a different number of individuals will contribute to different elements in W. To ease notation, this is left implicit in the above formula.

It has been tested whether the parameters are different for the two education groups

[10] Under some regularity conditions. See Hansen (1982) for a detailed exposition.

considered in the estimation. Given the asymptotic normal distribution of the EWMD estimator and the fact that the two samples are independent, the Wald statistic has been computed to test the joint hypothesis that all the parameters are equal in the two groups (group 1 and group 2):

$$\mathrm{X} = \left(\hat{\alpha}^1 - \hat{\alpha}^2\right)' \left(W^1 + W^2\right)\left(\hat{\alpha}^1 - \hat{\alpha}^2\right)$$

which is distributed as a chi-square with *n* (the dimension of the parameter vector) degrees of freedom.

MAPPING

1) permanent-transitory model

$$u_{it}^a = \mu_i + \varepsilon_{it}^a$$

where i.i.d. measurement error is captured by the transitory component.
Estimation of this model is feasible if one observes the variance of earnings and its covariance. The theoretical moments are:

$$\mathrm{var}(u_{it}^a) = \sigma_\mu^2 + \sigma_\varepsilon^2 \qquad \text{and}$$

$$\mathrm{cov}(u_{it}^a,\, u_{i,t-s}^{a-s}) = \sigma_\mu^2 \qquad s=2, 3, 4, \ldots$$

2) AR(1) model

$$u_{it}^a = \mu_i + z_{it}^a$$
$$z_{it}^a = \alpha\, z_{i,t-1}^{a-1} + \omega_{it}^a$$

Individuals start working at age $\underline{a}$ ($\underline{a} = 24$)
Moments are built as:

$$Var(u_{it}^a) = Var(\mu_i) + Var(z_{it}^a)$$

$$Var(\mu_i) = \sigma_\mu^2$$

$$Var(z_{it}^{\underline{a}}) = \sigma_{\underline{a}}^2$$

$$Var(z_{it}^{a}) = \alpha^2 Var(z_{it-1}^{a-1}) + \sigma_\omega^2 \qquad a > \underline{a}$$

$$Cov\left(u_{it}^{a}, u_{it-s}^{a-s}\right) = \sigma_\mu^2 + Cov\left(z_{it}^{a}, z_{it-s}^{a-s}\right) \qquad s\text{=2, 3, 4...}$$

$$Cov\left(z_{it}^{a}, z_{it-s}^{a-s}\right) = \alpha^s \, Var(z_{it-s}^{a-s}) \qquad s\text{=2, 3, 4...}$$

3) ARMA(1,1) model

$$u_{it}^{a} = \mu_i + z_{it}^{a}$$
$$z_{it}^{a} = \alpha \, z_{i,t-1}^{a-1} + \xi_{it}^{a} + \theta \xi_{it-1}^{a-1}$$

Moments are built as:

$$Var(u_{it}^{a}) = Var(\mu_i) + Var(z_{it}^{a})$$

$$Var(\mu_i) = \sigma_\mu^2$$

$$Var(z_{it}^{\underline{a}}) = \sigma_\xi^2$$

$$Var(z_{it}^{a}) = \alpha^2 Var(z_{it-1}^{a-1}) + \left(1 + \theta^2\right)\sigma_\xi^2 \qquad a > \underline{a}$$

$$Cov\left(u_{it}^{a}, u_{it-s}^{a-s}\right) = \sigma_\mu^2 + Cov\left(z_{it}^{a}, z_{it-s}^{a-s}\right)$$

$$Cov\left(z_{it}^{a}, z_{it-s}^{a-s}\right) = \alpha^s \, Var(z_{it-s}^{a-s}) + \alpha^{s-1}\theta \, \sigma_\xi^2 \qquad s\text{=2, 3, 4...}$$

SAMPLE MOMENTS

Sample moments have been built on the residuals of regressions of the logarithm of gross yearly earnings on an age polynomial and cohort and time dummies. Age, cohort and year effects cannot be separately identified without making some further assumptions, as they are linear combinations of one another. In the analysis it is therefore assumed that the time effects are orthogonal to a time trend and add up to zero.

In order to control for education, regressions are estimated separately for each education group (up to 5 years, 8 years, 13 years or 17+ years of education). To compute the sample moments, only two education groups have been considered (high school dropouts and high school and college graduates).

Five-year date-of-birth cohorts have been built, the younger cohort being born in 1963-1967, and the oldest in 1938-1942. The resulting cohorts are six and they are observed for 5 time intervals.

In the panel data set, individuals belonging to the younger cohort are observed at ages 24, 26, 28, 30 and 33. For these individuals it is therefore possible to compute 5 variances, 3 second-order covariances, one third covariance and so on. Other cohorts are treated similarly. In total there are 30 variances, 18 second order covariances, 6 lag three covariances, and so on.

Our papers can be downloaded at:

http://cerp.unito.it

CeRP Working Paper Series:

N° 1/00	Guido Menzio	Opting Out of Social Security over the Life Cycle
N° 2/00	Pier Marco Ferraresi Elsa Fornero	Social Security Transition in Italy: Costs, Distorsions and (some) Possible Correction
N° 3/00	Emanuele Baldacci Luca Inglese	Le caratteristiche socio economiche dei pensionati in Italia. Analisi della distribuzione dei redditi da pensione (*only available in the Italian version*)
N° 4/01	Peter Diamond	Towards an Optimal Social Security Design
N° 5/01	Vincenzo Andrietti	Occupational Pensions and Interfirm Job Mobility in the European Union. Evidence from the ECHP Survey
N° 6/01	Flavia Coda Moscarola	The Effects of Immigration Inflows on the Sustainability of the Italian Welfare State
N° 7/01	Margherita Borella	The Error Structure of Earnings: an Analysis on Italian Longitudinal Data