Supplement to "Grade retention and unobserved heterogeneity"

(*Quantitative Economics*, Vol. 7, No. 3, November 2016, 781–820)

Robert J. Gary-Bobo
CREST, ENSAE

Marion Goussé
Department of Economics, Université Laval

Jean-Marc Robin
Department of Economics, Sciences Po and University College London

We study the treatment effect of grade retention using a panel of French junior high-school students, taking unobserved heterogeneity and the endogeneity of grade repetitions into account. We specify a multistage model of human-capital accumulation with a finite number of types representing unobserved individual characteristics. Class-size and latent student-performance indices are assumed to follow finite mixtures of normal distributions. Grade retention may increase or decrease the student's knowledge capital in a type-dependent way. Our estimation results show that the average treatment effect on the treated (ATT) of grade retention on test scores is positive but small at the end of grade 9. Treatment effects are heterogeneous: we find that the ATT of grade retention is higher for the weakest students. We also show that class size is endogenous and tends to increase with unobserved student ability. The average treatment effect of grade retention is negative, again with the exception of the weakest group of students. Grade repetitions reduce the probability of access to grade 9 of all student types.

Keywords. Secondary education, grade retention, unobserved heterogeneity, finite mixtures of normal distributions, treatment effects, class-size effects.

Robert J. Gary-Bobo: robert.gary-bobo@ensae.fr
Marion Goussé: marion.gousse@gmail.com
Jean-Marc Robin: jeanmarc.robin@sciencespo.fr
We thank Costas Meghir, Petra Todd, and three anonymous referees for useful comments. Robin gratefully acknowledges financial support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice, Grant RES-589-28-0001, and from the European Research Council (ERC), Grant ERC-2010-AdG-269693-WASP. Gary-Bobo's research is supported by Labex Ecodec (ANR-11-LABX-0047) and Investissements d'Avenir (ANR-11-IDEX-0003). Robin and Gary-Bobo gratefully acknowledge support from the Agence Nationale de la Recherche (ANR) white project "Econométrie des redoublements."

DOI: 10.3982/QE524

Appendix B: 3SLS estimation of a linear model

As a benchmark for the structural approach presented in the main paper, we provide here details on estimation of a standard linear model with five simultaneous equations explaining the final (grade-9) test scores Y_{1m} and Y_{1f} in math and French, respectively, grade repetition R, class size experienced in grade 9, denoted N, and class size in grade 6, denoted N_0. These variables are explained as a function of a long list of controls X, including parental occupation and parental education; the initial (grade-6) test scores in math and French, respectively, denoted Y_{0m}, Y_{0f}; if retention intervened in primary school, denoted R_0, the instrument for grade repetition is the semester of birth (3rd or 4th quarter), denoted Q; the instruments for class size are the theoretical class size (à la Angrist–Lavy), denoted Z_1, and total school enrollment, denoted T_1 in grade 9; finally, the same variables are denoted Z_0 and T_0 when measured in grade 6. We estimate a simplified model that does not explain transitions from grade 6 to grade 7, from grade 7 to grade 8, and from grade 8 to grade 9. The model is specified as

$$Y_{1m} = a_m R + X_1 b_m + c_{mm} Y_{0m} + c_{mf} Y_{0f} + d_m R_0 + u_m, \tag{S1}$$

$$Y_{1f} = a_f R + X_1 b_f + c_{fm} Y_{0m} + c_{ff} Y_{0f} + d_f R_0 + u_f, \tag{S2}$$

$$N_1 = a_n R + \alpha_{t1} T_1 + \alpha_{z1} Z_1 + X_1 \beta_1 + \gamma_{1m} Y_{0m} + \gamma_{1f} Y_{0f} + \delta_1 R_0 + w_1, \tag{S3}$$

$$R = a_r Q + X_1 b_r + c_{rm} Y_{0m} + c_{rf} Y_{0f} + d_r R_0 + v, \tag{S4}$$

$$N_0 = \alpha_{t0} T_0 + \alpha_{z0} Z_0 + X_0 \beta_0 + \gamma_{0m} Y_{0m} + \gamma_{0f} Y_{0f} + \delta_0 R_0 + w_0. \tag{S5}$$

The parameters are $(a, b, c, d, \alpha, \beta, \gamma, \delta)$, with subscripts m standing for math, f for French, and r for repetition. The basic assumption justifying this model is that the student's underlying talent or type is captured by entry test scores and retention in primary school (Y_{0m}, Y_{0f}, R_0). We assume that error terms (u_m, u_f, v, w_1, w_0) are independent from $(Q, R_0, T_i, X_i, Y_{0m}, Y_{0f}, Z_i)$, $i = 0, 1$. In essence, the error terms represent other unobserved characteristics of individuals that are assumed to be orthogonal, not only to instruments and family-background controls, but also to the predetermined performance measures (R_0, Y_{0m}, Y_{0f}). Table S1 gives the results of a particular variant of this model.

Table S1 gives the key coefficients of the five equations. Table S2, which is the continuation of Table S1, displays the coefficients of a number of family-background controls, to allow for an easy comparison with the results presented in Section 6. Table S3 reports the ATT and ATE values obtained when the coefficients of the linear model are used. Table S4 gives the correlation matrix of error terms (u_m, u_f, w_1, v, w_0) as obtained from the 3SLS estimation of the covariance matrix of residuals. We tried several variants of this type of model and found that most of the key coefficients are significant: the impact of higher entry test scores is positive on final scores, negative on grade repetition, and positive on class size, because class size is used as a remedial instrument—weaker students being assigned to smaller classes. We even find a significant impact of class size on final

TABLE S1. The 3SLS estimates of the linear model.

Panel A Dependent Variable	(1) Y_{1m}	(2) Y_{1f}	(3) N_1	(4) R	(5) N_0
Repetition (R)	11.85***	−7.715**	3.498**		
	(3.755)	(3.489)	(1.470)		
Initial score in math	0.663***	0.175***	0.0662***	−0.0116***	0.0288***
	(0.0442)	(0.0410)	(0.0170)	(0.000519)	(0.00397)
Initial score in French	0.255***	0.359***	0.0704***	−0.00990***	0.0178***
	(0.0389)	(0.0361)	(0.0148)	(0.000513)	(0.00400)
Retention in primary school	−1.054***	−2.386***	−0.201*	−0.0541***	−0.380***
	(0.302)	(0.275)	(0.118)	(0.00953)	(0.0743)
Class size in grade 9 (N_1)	−0.453***	−0.302***			
	(0.0863)	(0.0797)			
Theor. class size grade 9			0.282***		
			(0.0193)		
Total school enrollment grade 9			0.00196***		
			(0.000209)		
Class size grade 6 (N_0)				0.0184***	
				(0.00448)	
Born in 2nd semester				0.0192***	
				(0.00537)	
Theor. class size grade 6					0.286***
					(0.0200)
Total school enrollment grade 6					0.000972***
					(0.000178)
Constant	11.20**	33.00***	9.818***	0.831***	15.71***
	(4.500)	(4.153)	(1.877)	(0.0991)	(0.399)
Observations	11,648	11,648	11,648	11,648	11,648
R-squared	0.178	0.415	0.062	0.198	0.161

Note: This table and the following one give the estimation results for the five-equations linear model (S1)–(S5), using the 3SLS method. Specific IVs are the semester of birth (to instrument grade retention), total school enrollment, and theoretical class size (Angrist–Lavy's instrument) for class size. There are other controls whose coefficients are not reported; in the five equations we control for the number of inhabitants in the jurisdiction (i.e., the commune). There are six dummies indicating intervals of urban size, a dummy for inhabitants of the Paris area, and another dummy for areas with special school subsidies (ZEP schools). Standard errors are given in parentheses. The asterisks ***, **, and * indicate significance at the 1, 5, and 10% levels, respectively.

grades that has the right (negative) sign. To sum up, the key coefficients are significant, with the expected sign, except the main parameter of the study, namely, the impact of the grade repetition dummy R on final scores. The estimates of (a_m, a_f) are typically unstable: their magnitude seems too large, they depend on the list of controls X_0 and X_1, and they also depend on the list of exclusions. In Table S1, the signs of a_m and a_f are also different, in sharp contrast with the results of Table 4 above, and these signs can change or one of the coefficients can become nonsignificant in some variants. We probably do not have the right instrument for grade repetition; it is unclear which local treatment effect is captured by Q.

TABLE S2. The 3SLS estimates: Impact of family background.

Panel B Dependent Variable	(1) Y_{1m}	(2) Y_{1f}	(3) N_1	(4) R	(5) N_0
Female	−1.553***	−3.830***	−0.317***	0.0565***	0.0253
	(0.268)	(0.245)	(0.103)	(0.00688)	(0.0544)
Mother's education, ref.: no educ.					
Middle-school certificate	0.293	−0.120	−0.0832	0.000193	0.0741
	(0.247)	(0.221)	(0.0938)	(0.0105)	(0.0829)
Vocational certificate	0.0654	0.0111	−0.279***	0.0164	0.0944
	(0.271)	(0.242)	(0.102)	(0.0111)	(0.0878)
High-school degree	1.810***	0.959***	0.114	−0.0157	−0.0510
	(0.305)	(0.273)	(0.116)	(0.0126)	(0.100)
2 years of college	2.578***	1.231***	0.473***	−0.0398***	0.179*
	(0.355)	(0.320)	(0.133)	(0.0136)	(0.108)
4 years of college and more	3.068***	2.521***	0.710***	−0.0269*	0.290**
	(0.398)	(0.357)	(0.148)	(0.0162)	(0.127)
Father's occupation, ref.: blue collar					
Farmer	2.623***	−0.303	0.201	−0.0432**	−0.199
	(0.512)	(0.460)	(0.195)	(0.0201)	(0.159)
Self-employed,	0.155	−0.676**	0.450***	−0.0228*	0.354***
Head of own business	(0.298)	(0.267)	(0.112)	(0.0122)	(0.0959)
Executive, professional	2.003***	0.739***	0.514***	−0.0320***	0.302***
	(0.293)	(0.264)	(0.109)	(0.0115)	(0.0899)
Intermediate profession	1.243***	0.549**	0.251***	−0.0189*	0.142*
	(0.248)	(0.222)	(0.0934)	(0.0101)	(0.0797)
White collar employee	0.314	0.462*	0.0576	−0.00888	−0.0163
	(0.278)	(0.248)	(0.105)	(0.0116)	(0.0922)
Unemployed	−1.861***	−1.257***	−0.358**	0.0215	0.162
	(0.405)	(0.363)	(0.153)	(0.0165)	(0.131)
Three children family	−0.0197	0.244	−0.131*	0.0221***	−0.107*
	(0.193)	(0.173)	(0.0734)	(0.00756)	(0.0597)
More than three children family	−0.627***	−0.341	−0.104	0.00379	−0.0262
	(0.238)	(0.212)	(0.0901)	(0.0100)	(0.0794)

Note: This table is just the continuation of Table S1. The same remarks apply. The number of observations is $N = 11{,}648$. Standard errors are given in parentheses. The asterisks ***, **, and * indicate significance at the 1, 5, and 10% levels, respectively.

TABLE S3. The ATE and ATT computed with 3SLS estimates.

ATE	
Math	8.65
French	−9.85
ATT	
Math	10.27
French	−8.77

TABLE S4. Correlation matrix of error terms (3SLS).

	Y_m	Y_f	N_1	R	N_0
Y_m	1				
Y_f	0.37	1			
N_1	0.25	0.02	1		
R	−0.4	0.24	−0.34	1	
N_0	0.01	0.02	0.15	−0.12	1

TABLE S5. Probabilities of types (variant).

	Group 1	Group 2	Group 3	Group 4
Support point	37.95	50.27	51.38	59.57
	(0.26)	(0.65)	(0.3)	(0.13)
Prior probability	23.54%	21.66%	28.91%	25.89%
	(0.72)	(4.17)	(4.17)	(0.84)

Note: The probabilities are the prior probabilities of types, computed as indicated by equations (20) and (21). The support of types is conventionally obtained as the list of the estimated coefficients of group dummies in a regression of the entry test score in mathematics on group dummies only. This shows that group 2 and group 3 have approximately the same support point. In Table S6, additional controls are included in the regressions, but this changes the values of support points only slightly.

APPENDIX C: VARIANT OF THE MODEL WITH UNOBSERVED HETEROGENEITY

To check the robustness of our approach, we reestimated the model of Sections 4–6 using the same EM algorithm, the same specification, and the same number of latent types, except that we introduced a limited number of controls in the equations. This was done to see if the introduction of family-background controls can alter the location and probabilities of latent student types and ultimately, if the ATTs and ATEs of grade retention would also change. We added four variables describing the socioeconomic environment of the student: (i) the student gender dummy; (ii) a dummy indicating if the mother is a college graduate (with the equivalent of more than 4 years of college); (iii) a dummy indicating if the father is an executive with a higher education or a professional (lawyers, engineers, doctors, teachers, ...); (iv) a dummy indicating grade repetition in primary school. These variables are added as controls in all the equations. In addition, the quarter-of-birth dummies have been added in all the promotion–retention, Probit, and ordered Probit equations. We then ran the EM algorithm on this modified version. Standard errors are obtained by bootstrapping.

Table S5 gives the new ex ante probabilities of types and the new support points for these probabilities, as obtained with the variant. We observe that types 2 and 3 are now difficult to distinguish. It seems that the model has only three types. This can be seen more precisely on Table S6, which gives the coefficients of group indicators in equations (1)–(4) and (15)–(18) (this table is the equivalent of Table 10 for the variant under study). In each of the columns of Table S6, the coefficients for group 3 are always very close to the corresponding coefficient for group 2. In Table S5, we see that the prior probabilities of group 2 and group 3 are estimated with less precision than the other two probabilities.

Table S6. Regression of test scores on group indicators and class size (variant).

	Score in Math			Score in French		
		Final			Final	
	Initial	Nonrepeaters	Repeaters	Initial	Nonrepeaters	Repeaters
Class size		−0.004			−0.066	
		(0.064)			(0.052)	
Class size rep.			−0.082			−0.093
			(0.091)			(0.074)
Group 2	11.60***	9.99***	5.59***	10.52***	8.93***	5.32***
	(0.52)	(1.38)	(1.31)	(0.70)	(0.69)	(1.06)
Group 3	12.59***	8.45***	5.09***	11.86***	8.45***	5.22***
	(0.29)	(1.34)	(1.12)	(0.51)	(0.79)	(1.04)
Group 4	20.75***	20.07***	10.98***	19.60***	19.83***	11.26***
	(0.22)	(0.39)	(1.39)	(0.20)	(0.34)	(1.16)
Constant	38.17***	39.85***	42.80***	41.05***	43.81***	45.29***
	(0.27)	(1.62)	(2.26)	(0.25)	(1.32)	(2.01)
R^2	67.71%	58.13%	23.12%	65.42%	59.51%	23.59%

Note: There are $N = 12{,}916$ observations. Standard errors are given in parentheses. The asterisks ***, **, and * indicate significance at the 1, 5, and 10% levels, respectively.

Table S7. Average treatment effect (variant).

	Math		French	
	ATE	ATT	ATE	ATT
Group 1	1.13	1.24	1.33	1.5
	(0.4)	(0.42)	(0.42)	(0.43)
Group 2	−3.13	−2.77	−2.2	−1.6
	(1.53)	(1.56)	(1.15)	(1.18)
Group 3	−2.25	−1.79	−1.9	−1.18
	(1.43)	(1.46)	(1.24)	(1.24)
Group 4	−7.91	−7.03	−7.23	−6
	(1.33)	(1.24)	(1.16)	(1.09)
All	−3.3	−0.66	−2.78	−0.1
	(0.37)	(0.23)	(0.3)	(0.22)

Note: Standard deviations, in parentheses, were obtained by bootstrapping.

But apart from these differences, we see in Table S7 that the end results are more or less the same with the variant: ATE and ATT are positive only for group 1; the last line of Table S7 shows that the overall ATT is negative but small, barely significant, while the ATE is clearly negative. This shows that our results are robust to this kind of change in the specification.

It is also interesting to study the relationship between the old types, the new types, and the family-background controls. This is done in Table S8. In the model and its variant, each individual is characterized by a probability of belonging to group k for each $k = 1, \dots, 4$. The idea is to regress the posterior probability of belonging to group k in the "old" model, denoted $p_{ik} = \mathbb{P}(G_{ik} = 1|X, Y, Z)$, on the probabilities of belonging to group k' in the "new model," $k' = 1, \dots, 4$, denoted $p'_{ik} = \mathbb{P}(G'_{ik} = 1|X, Y, Z)$, with the controls for family background introduced in the new model. This yields the regressions, for all groups $j = 1, \dots, 4$,

$$p_{ij} = \sum_k \alpha_{jk} p'_{ik} + X_i \beta + u_i.$$

Assuming that the random error term u_i has a zero mean, is independent of controls X_i, and is independent of posterior probabilities p'_{ik}, we find that

$$\mathbb{E}\big(p_{ij}|p'_{ik} = 1\big) = \alpha_{jk} + X_i \beta.$$

This provides a way to interpret the coefficients in the upper half of Table S8. These coefficients are entries in the 4×4 matrix $A = (\alpha_{jk})$. In Table S8, the boldface coefficients are higher than 0.2; the other ones are comparatively much smaller. If the types in the old model were more or less the same as in the new model with controls added, the diagonal terms α_{kk} would be the only nonnegligible entries of matrix A. This is true only for the weakest and the strongest groups, that is, group 1 and group 4, respectively, which seem

TABLE S8. Regression of "old" individual posterior probabilities on "new" posterior probabilities.

Dependent Variables:	Prob Group 1	Prob Group 2	Prob Group 3	Prob Group 4
Prob group 1′	**0.538*****	**0.393*****	0.0344***	0.0346***
	(0.00495)	(0.00856)	(0.00844)	(0.00507)
Prob group 2′	−0.00523	**0.509*****	**0.438*****	0.0584***
	(0.00537)	(0.00928)	(0.00914)	(0.00550)
Prob group 3′	−0.0422***	**0.392*****	**0.617*****	0.0326***
	(0.00454)	(0.00785)	(0.00774)	(0.00466)
Prob group 4′	−0.0377***	0.000239	**0.322*****	**0.716*****
	(0.00471)	(0.00814)	(0.00802)	(0.00482)
Female	0.0395***	0.0328***	−0.000341	−0.0719***
	(0.00406)	(0.00701)	(0.00691)	(0.00416)
Father is an executive	−0.0681***	−0.138***	0.0465***	0.159***
	(0.00585)	(0.0101)	(0.00996)	(0.00599)
Mother: college graduate	−0.0276***	−0.111***	−0.0380**	0.176***
	(0.00880)	(0.0152)	(0.0150)	(0.00901)
Retention in	0.199***	0.0438***	−0.165***	−0.0777***
primary school	(0.00511)	(0.00884)	(0.00871)	(0.00524)
R^2	0.639	0.464	0.518	0.708

Note: The posterior probabilities of belonging to a group, in the model presented in Sections 4–6 (called the old model), are regressed on the posterior probabilities obtained in the variant presented in Appendix C (the new model) for each of the four groups plus the controls introduced in this variant. The estimation method here is OLS. The coefficients greater than 0.2 are set in boldface; most of the other coefficients are smaller than 0.05. This shows that old group 1 and old group 4 are close to new group 1 and new group 4, respectively. These groups seem well identified. There are $N = 12{,}916$ observations. Standard errors are given in parentheses. The asterisks ***, **, and * indicate significance at the 1, 5, and 10% levels, respectively.

well identified in both variants of the model. But the old group 3 is recruiting individuals from new groups 2, 3, and 4, while the old group 2 is filled with individuals who belong to new groups 1, 2, and 3.

In addition, Table S8 gives indications about the role played by family background and gender in the determination of the distribution of individuals in the old groups. The coefficients β are significantly different from zero with the “expected” sign. Educated mothers and well-to-do fathers increase (resp. decrease) the probability of belonging to group 4 (resp. group 1). Retention in elementary school increases the probability of belonging to group 1. From other regression results, for instance, the 3SLS estimates in Appendix B, we know that with the data at hand, female students have slightly lower scores and are a bit more likely to be held back; it is then not surprising to find that they are slightly more likely to belong to old groups 1 and 2. We conclude that there is a form of robustness of the classification of individuals provided by the model. The introduction of controls has made the identification of intermediate groups more difficult, since new groups 2 and 3 are very close, but the weakest and the strongest, that is, group 1 and group 4, are well identified in both variants.

Co-editor Petra E. Todd handled this manuscript.

Submitted December, 2014. Final version accepted January, 2016.