Inference on an Extended Roy Model, with an Application to Schooling Decisions in France Online Appendix

Xavier D'Haultfœuille * Arnaud Maurel†

October 4, 2012

Abstract

In this appendix we discuss some Monte Carlo simulations to evaluate the accuracy of our three-stage estimator in finite samples. We then provide some descriptives and details on the computation of the stream of earnings in the application, as well as additional results and robustness checks. Finally, the third section is devoted to the proofs of our results.

1 Monte Carlo simulations

In this section, we investigate the finite-sample performance of our semiparametric estimation procedure by simulating the following model with sector-specific variables, for four different sample sizes (namely $n = 500$, $n = 1,000$, $n = 2,000$ and $n = 10,000$):

$$\begin{aligned} Y_{0i} &= X_{2i}\beta_{02} + X_{3i}\beta_{03} + \eta_{0i} + \nu_{0i} \\ Y_{1i} &= X_{1i}\beta_{11} + X_{3i}\beta_{13} + \eta_{1i} + \nu_{1i} \\ D_i &= \mathbb{1}\{-\delta_0 + X_{1i}(\beta_{11} - \gamma_{01}) + X_{2i}(-\beta_{02} - \gamma_{02}) + X_{3i}(\beta_{13} - \beta_{03} - \gamma_{03}) + \eta_{1i} - \eta_{0i} > 0\}. \end{aligned}$$

The true values of the parameters are $\beta_{02} = \beta_{03} = 1$, $\beta_{11} = 2$, $\beta_{13} = 0.5$, $\gamma_{01} = -0.5$, $\gamma_{02} = 0.5$, $\gamma_{03} = -0.8$ and $\delta_0 = 0.8$, so that Assumption 3.1 is satisfied with $j_1 = 1$ and $j_2 = 2$. We simulate X_{1i} and X_{2i} independently and from a uniform distribution over $[0, 4]$, while X_{3i} is a discrete regressor drawn from a Bernoulli distribution

*CREST. E-mail address: xavier.dhaultfoeuille@ensae.fr.

†Duke University and IZA. E-mail address: apm16@duke.edu.

with parameter 0.5. We let $(\eta_{0i}, \eta_{1i})'$ be joint normal, with mean $(0,1)'$ and a variance Σ such that $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = 0.5$. $(\nu_{0i}, \nu_{1i})'$ are drawn from a heteroskedastic normal distribution, with zero mean and a conditional matrix variance $\Omega(X)$ such that $\Omega_{11}(X) = \exp(X_2/5)$, $\Omega_{22}(X) = \exp(X_1/5)$ and $\Omega_{12}(X) = \Omega_{21}(X) = 0.5\sqrt{\Omega_{11}(X)\Omega_{22}(X)}$.

We implement the three-stage estimation procedure detailed in Section 3 of the main text. We estimate in the first stage $\zeta_0 = (\beta_1 - \beta_0 - \gamma_0)/(\beta_{11} - \gamma_{01})$ by Klein & Spady's (1993) semiparametric efficient estimator, with an adaptive gaussian kernel and local smoothing. In the second stage, we implement Newey's (2009) method in order to estimate separately β_0 and β_1. The series estimator of the selection correction term was computed using the inverse Mills ratio transform and Legendre polynomials of order increasing with the sample size at a rate $n^{1/8}$ (see Newey, 2009, Equation (3.6) and Assumption (4.5) respectively).[1] Using Legendre polynomials instead of simple power series lead, theoretically speaking, to the same results but avoids numerical troubles due to quasi multicollinearity (see Newey, 2009). In the third stage, we finally implement our proposed estimators for δ_0 and γ_0 with the quartic kernel suggested in Section 3 of the main text and a bandwidth $h_n = 0.5\sigma(\widehat{U})n^{-1/7}$, where $\sigma(\widehat{U})$ is the estimated standard deviation of $\widehat{U}$. We choose the functions $h_1(x) = \Phi(\widehat{a}_0 + \widehat{a}_1 x)$ and $h_2(x) = xh_1(x) - \int_{\widehat{u}_0}^{x} \widehat{q}(u, \widehat{\zeta})du$ for the instruments, where $\Phi(.)$ denotes the normal cumulative distribution, $(\widehat{a}_0, \widehat{a}_1)$ is the probit estimator of D on $(1, \widehat{U})$ and $\widehat{u}_0$ is the sample minimum of $\widehat{U}$.[2] Finally, no trimming was performed since it did not improve the accuracy of the estimators in our setting.

The performance of the estimators are summarized in Panel A of Table 1, which reports for each parameter its bias, standard deviation and root mean squared error (RMSE). The non-pecuniary components γ_0 and δ_0 are less precisely estimated and the corresponding estimators display a larger bias than that of the outcome equations parameters, β_0 and β_1. Basically, this stems from the sequential structure of the proposed estimation procedure, and from the fact that, in this specification, no exclusion restriction is used to identify γ_0 and δ_0. Despite this, the estimators of γ_0 and δ_0 are accurate enough for being able to reject in most simulation samples the hypothesis that $\delta_0 = 0$ or $\gamma_{03} = 0$ for $n = 2,000$ and $n = 10,000$, as well as the hypothesis that $\gamma_{02} = 0$ for

[1]Namely, polynomials of order 6 are used for sample sizes $n = 500$ and $n = 1,000$, order 7 for $n = 2,000$ and 8 for $n = 10,000$.

[2]For the sake of simplicity, we suppose in Section 3 of the main text that the functions $h(.)$ are known to the econometrician. Assuming alternatively that these functions have to be estimated, as is the case here, does not affect the root-n consistency and asymptotic normality of the estimators.

$n = 10,000$. Besides, even for small sample sizes, these estimators display a negligible bias, compared to their standard deviation, which is reassuring for conducting valid inference.

We also investigate the effect of using an exclusion restriction on the non-pecuniary component on the finite-sample performances of the estimators. For that purpose, we consider the same specification as previously with the exception that $\gamma_{01} = 0$, and compare estimates obtained when this restriction is known by the econometrician and when it is not. As explained in the main text, we can recover γ_0 in the former case with the first step estimates alone, and use Equation (3.7) in the main text to estimate δ_0 only. The properties of the unconstrained and constrained estimators are displayed respectively in Panel B and C of Table 1. Using an exclusion restriction on the non-pecuniary component leads to a substantial improvement in the performances of the estimators of γ_0, for all sample sizes. Notably, the standard error for γ_{02} decreases by about 80% between the two specifications. The performance of the estimator for δ_0, which is still estimated in a third step, is very similar to the unconstrained specification. Overall, it appears from these Monte Carlo simulations that the constrained estimator should be preferred in the presence of an exclusion restriction on the non-pecuniary component. Importantly, whether such an exclusion restriction is valid can be directly tested in the data after estimating the unconstrained model.

n	Coeff.	Panel A			Panel B			Panel C		
		Bias	Std dev	RMSE	Bias	Std dev	RMSE	Bias	Std dev	RMSE
500	β_{02}	-0.009	0.174	0.174	-0.006	0.161	0.161	-0.006	0.161	0.161
	β_{03}	-0.011	0.288	0.288	-0.003	0.259	0.259	-0.003	0.259	0.259
	β_{11}	0.003	0.136	0.136	0.004	0.124	0.124	0.004	0.124	0.124
	β_{13}	-0.001	0.172	0.172	0.000	0.187	0.187	0.000	0.187	0.187
	γ_{01}	0.083	1.781	1.783	-0.017	1.261	1.261	*(not estimated)*		
	γ_{02}	-0.106	1.033	1.038	-0.021	0.923	0.923	-0.032	0.207	0.209
	γ_{03}	0.035	0.475	0.476	0.012	0.420	0.420	0.020	0.356	0.357
	δ_0	0.090	0.747	0.753	0.036	0.673	0.674	0.037	0.642	0.643
1,000	β_{02}	-0.009	0.124	0.124	-0.006	0.110	0.110	-0.006	0.110	0.110
	β_{03}	-0.003	0.206	0.206	0.002	0.180	0.180	0.002	0.180	0.180
	β_{11}	0.005	0.089	0.089	-0.001	0.086	0.086	-0.001	0.086	0.086
	β_{13}	0.001	0.122	0.122	0.008	0.127	0.127	0.008	0.127	0.127
	γ_{01}	-0.021	1.201	1.201	0.063	0.879	0.881	*(not estimated)*		
	γ_{02}	-0.012	0.695	0.695	-0.067	0.651	0.655	-0.019	0.140	0.141
	γ_{03}	0.020	0.323	0.323	0.013	0.297	0.297	0.006	0.251	0.251
	δ_0	0.022	0.510	0.510	0.037	0.454	0.456	0.031	0.446	0.447
2,000	β_{02}	-0.005	0.090	0.090	0.000	0.080	0.080	0.000	0.080	0.080
	β_{03}	-0.001	0.136	0.136	0.003	0.125	0.125	0.003	0.125	0.125
	β_{11}	0.001	0.067	0.067	0.002	0.062	0.062	0.002	0.062	0.062
	β_{13}	-0.001	0.091	0.091	0.002	0.091	0.091	0.002	0.091	0.091
	γ_{01}	0.047	0.883	0.884	0.031	0.628	0.629	*(not estimated)*		
	γ_{02}	-0.039	0.517	0.519	-0.036	0.466	0.467	-0.014	0.100	0.101
	γ_{03}	0.021	0.233	0.234	0.006	0.210	0.210	0.004	0.184	0.184
	δ_0	0.012	0.360	0.360	0.023	0.332	0.333	0.023	0.331	0.332
10,000	β_{02}	0.003	0.039	0.039	-0.002	0.034	0.034	-0.003	0.036	0.036
	β_{03}	-0.005	0.067	0.067	0.001	0.057	0.057	0.002	0.057	0.058
	β_{11}	-0.003	0.028	0.028	0.001	0.026	0.027	0.000	0.027	0.027
	β_{13}	-0.002	0.039	0.039	-0.004	0.041	0.041	-0.001	0.041	0.041
	γ_{01}	0.084	0.385	0.394	0.015	0.276	0.276	*(not estimated)*		
	γ_{02}	-0.024	0.232	0.233	-0.010	0.204	0.204	0.000	0.043	0.043
	γ_{03}	0.012	0.104	0.105	-0.001	0.088	0.088	-0.005	0.079	0.079
	δ_0	-0.049	0.153	0.161	-0.004	0.136	0.137	0.000	0.141	0.142

Note: Panel A corresponds to the unconstrained model, while in Panel B and Panel C, $\gamma_{01} = 0$. In Panel B we suppose that the econometrician ignores this restriction, so that (δ_0, γ_0) are estimated using Equation (3.5) in the main text. In panel C, the econometrician knows it, and estimates are based on Equation (3.7) in the main text. The results were obtained with 1,000 simulations for each sample size.

Table 1: Monte Carlo simulations

2 Higher education attendance in France: supplementary material

We report in Table 2 below some descriptive statistics for the subsample of interest, according to higher education attendance.

Variable	Higher education attendees Mean	Higher education attendees Std. dev.	High school level Mean	High school level Std. dev.
Initial monthly log wage (1992 French Francs)	8.75	0.44	8.50	0.39
Secondary schooling track				
L (Humanities)	0.15	0.36	0.04	0.19
ES (Economics and Social Sciences)	0.17	0.38	0.04	0.19
S (Sciences)	0.32	0.47	0.06	0.23
Vocational	0.04	0.20	0.66	0.47
Technical	0.32	0.46	0.21	0.41
Born abroad	0.02	0.16	0.02	0.15
Father born abroad	0.11	0.32	0.11	0.32
Mother born abroad	0.10	0.31	0.10	0.30
Entering the labor market in 1992	0.46	0.50	0.51	0.50
Entering the labor market in 1998	0.54	0.50	0.49	0.50
Male	0.47	0.5	0.49	0.50
Father's profession				
Farmer	0.06	0.25	0.08	0.27
Tradesman	0.11	0.31	0.11	0.32
Executive	0.26	0.44	0.10	0.30
Intermediate occupation	0.12	0.32	0.09	0.29
Blue collar	0.17	0.38	0.30	0.46
White collar	0.21	0.41	0.25	0.44
Other	0.06	0.24	0.06	0.24
Age in 6^{th} grade				
≤ 10	0.10	0.29	0.03	0.17
11	0.84	0.37	0.72	0.45
≥ 12	0.07	0.25	0.25	0.43
Paris region	0.16	0.36	0.12	0.32
Number of higher education years	2.82	1.45	/	/
Dropout rate	0.16	0.37	/	/
Number of observations	19,143		5,082	

Table 2: Descriptive statistics.

2.1 Computation of the streams of earnings

For each alternative, the discounted streams of log-earnings are set equal to

$$Y_k^* = \sum_{t=t_{0,k}}^{t_{0,k}+A} \tau^t y_{k,t},$$

where $y_{k,t}$ denotes the flow of log-earnings received during year t, τ denotes the annual discount factor and A is the duration of active life. We account for the opportunity costs incurred when entering higher education by allowing the year of entry into the labor market ($t_{0,k}$) to vary according to the schooling choice. For a given year t, the variable $y_{k,t}$ is either set equal to the log-wage w_t earned during this period if the individual is employed at that time, or to the unemployment log-benefits b_t if the latter is unemployed. We set the replacement rate equal to 0.7 as often done in the literature.

We do not observe incomes during the whole life cycle in our data, so that we cannot compute $Y^* = DY_1^* + (1-D)Y_0^*$. Still, we can recover an expectation of this stream of income under additional assumptions on income dynamics. We suppose here that

$$y_{k,t} = \rho_k \mathbb{1}\{t - t_{0,k} + 1 \leq B\} + y_{k,t-1} + \nu_{k,t}, \tag{2.1}$$

where ρ_k denotes the alternative k-specific return to experience and $\nu_{k,t}$ is an alternative k-specific unobserved individual productivity term which is assumed to be independently and identically distributed over time, with mean zero. We introduce the dummy $\mathbb{1}\{t - t_{0,k} + 1 \leq B\}$ to account for non significant marginal returns to experience after B years of work (see, e.g., Kuruscu, 2006, for a similar assumption on wage growth). We also suppose that $\nu_{k,t}$ is independent of D, so that ρ_k is simply identified by $\rho_k = E(y_{k,t} - y_{k,t-1} | D = k)$, for $t \leq B + t_{0,k} - 1$.

Now, let $\widetilde{\tau}_k = \tau^{t_{0,k}} \left(\frac{1-\tau^{A+1}}{1-\tau}\right)$, $C_k = \tau^{t_{0,k}} \left(\frac{\tau}{(1-\tau)^2}\right) \left(1 - \tau^B + B\tau^{A+1}(\tau - 1)\right)$ and

$$Y_k = \widetilde{\tau}_k y_{D,t_{0,D}} + \rho_k C_k.$$

Because $\widetilde{\tau}_D$, C_D and ρ_D are identified for given τ, A and B, we can identify $Y = DY_1 + (1-D)Y_0$. Moreover, by Equation (2.1), $Y_k = E(Y_k^* | X, \eta_0, \eta_1, \nu_{k,t_{0,k}})$, which in turn implies that $E(Y_k | X, \eta_0, \eta_1) = E(Y_k^* | X, \eta_0, \eta_1)$. In other terms, the model may be written in terms of Y_k instead of Y_k^*, and our identification strategy applies with Y instead of the unobserved variable Y^*.

In practice, we set $\tau = 0.95$, $A = 45$ years, $B = 25$ years and estimate ρ_0 and ρ_1 to be respectively 0.025 and 0.042. These estimates were obtained by regressing $y_{k,t_{0,k}+T_k} - y_{k,t_{0,k}}$ on the number of years T_k for which the income is observed, on the subsample satisfying $D = k$.

2.2 Additional results

The first step estimates of $(\zeta, \beta_0, \beta_1)$ are displayed in Table 3. Overall, the results for β_0 and β_1 display a quite similar pattern. In particular, the local average income variables that we use as sector-specific variables have a strong positive effect, significant at the 1% level, on earnings.[3] Similarly, individuals entering the labor market in 1998 (relative to 1992) have very significantly higher earnings, reflecting the business cycle. However, some characteristics only affect the earnings of high school graduates or higher education attendees. This is in particular the case of gender, with high school male graduates earning significantly more than females. This is also the case of vocational secondary schooling tracks relative to technical tracks, which are positively related to earnings for high school graduates, while this is only true for male higher education attendees. Conversely, parental profession affects more significantly the earnings of higher education attendees than high school graduates, with negative signs associated with inactive, deceased or unemployed mother (referred to as "Other" in the table), relative to white collar professions. Similarly, higher education attendees whose mother is employed in an agricultural profession also earn significantly less.

[3] These local labor market variables were constructed by taking the average log-wages in the departement of residence at the time of entry into junior high school, weighted by the local rates of employment, over a 5-year time span centered respectively in 1992 or in 1998.

Variables	ζ	β_0	β_1
Local average income			
Higher education graduates	1.541*** (0.087)	0	0.019*** (0.004)
High school graduates	-1 (0)	0.022*** (0.004)	0
Secondary schooling track			
L	9.348*** (0.452)	-0.07* (0.039)	-0.011 (0.025)
ES	9.899*** (0.416)	-0.043 (0.04)	-0.002 (0.027)
S	10.133*** (0.426)	-0.055 (0.042)	-0.012 (0.026)
Vocational	-29.131*** (0.488)	0.247** (0.106)	-0.086 (0.094)
Technical	*Ref.*	*Ref.*	*Ref.*
Born abroad	1.727*** (0.46)	-0.006 (0.017)	0.000 (0.010)
Father born abroad	1.26** (0.451)	-0.011 (0.009)	0.011* (0.006)
Mother born abroad	1.591*** (0.464)	-0.018* (0.011)	0.007 (0.007)
Entering the labor market in 1998 (relative to 1992)	9.133*** (0.447)	0.097*** (0.035)	0.173*** (0.024)
Male	-0.298 (0.401)	0.043*** (0.008)	-0.001 (0.003)
Father's profession			
Farmer	2.291*** (0.434)	-0.012 (0.012)	0.014 (0.009)
Tradesman	1.289*** (0.43)	-0.008 (0.009)	-0.005 (0.005)
Executive	3.897*** (0.422)	-0.025 (0.016)	0.005 (0.011)
Intermediate occupation	1.799*** (0.457)	0.000 (0.009)	0.004 (0.007)
Blue collar	-0.49 (0.418)	0.008 (0.006)	-0.007 (0.004)
Other	1.309*** (0.432)	-0.013 (0.009)	-0.008 (0.006)
White collar	*Ref.*	*Ref.*	*Ref.*
Mother's profession			
Farmer	-6.343*** (0.51)	0.042* (0.025)	-0.038** (0.018)
Tradesman	-0.328 (0.488)	0.008 (0.01)	-0.002 (0.006)
Executive	1.279*** (0.469)	-0.01 (0.012)	-0.006 (0.006)
Intermediate occupation	0.899* (0.489)	-0.001 (0.009)	-0.001 (0.006)
Blue collar	-1.075** (0.438)	0.006 (0.008)	0.002 (0.006)
Other	-0.315 (0.411)	-0.003 (0.006)	-0.019*** (0.004)
White collar	*Ref.*	*Ref.*	*Ref.*
Age in 6th grade			
≤ 10	3.825*** (0.465)	-0.028 (0.017)	0.007 (0.01)
11	*Ref.*	*Ref.*	*Ref.*
≥ 12	-5.07*** (0.425)	0.035* (0.018)	-0.019 (0.013)
Paris region	1.181*** (0.453)	0.003 (0.012)	-0.002 (0.004)
Vocational $\times$ ...			
Entering the labor market in 1998	-1.012** (0.499)	-0.033* (0.018)	-0.021 (0.015)
Male	1.622*** (0.477)	-0.016 (0.01)	0.022** (0.01)
Paris region	-4.402*** (0.521)	0.023 (0.022)	-0.013 (0.018)

Standard errors, presented in parentheses, were computed by bootstrap with 200 bootstrap sample replicates. Significance levels: *** (1%), ** (5%) and * (10%).

Table 3: First step estimates.

2.3 Robustness to alternative computations of the streams of earnings

Finally, we also investigate the sensitivity of our results to the way the streams of earnings are computed. We reestimate the model with $\tau = 0.97$ instead of $\tau = 0.95$ (as, e.g., Carneiro et al., 2003), and $B = 30$ instead of $B = 25$. Results are displayed respectively in Panel 3 and 4 of Table 4 (Panel 1 and 2 corresponding to the robustness checks described in the main text). Once more, non-pecuniary components estimates are robust to this change. Standard errors, and thus the significance of some parameters, are slightly more affected by the specification choice. We also display in Figure 1 the estimate of the distribution of the *ex ante* returns to education with these alternative specifications.[4] Returns with $B = 30$ are nearly indistinguishable from the ones with $B = 25$. The distribution corresponding to $\tau = 0.97$ slightly dominates them, but remains within the confidence interval of the baseline specification. Finally, we also estimate the streams of earnings where people are aware of their own annual increase ρ_i of log-earnings, instead of just anticipating an average increase. We estimate ρ_i by OLS and compute the corresponding streams of earnings. Because of large errors on the estimated ρ_i and the sample reduction (ρ_i can be estimated only when at least two wages are reported, leaving us with only 9,364 individuals), no coefficient is significant anymore. The signs of γ remain however the same. Hence, our results are overall robust to alternative computations of Y.

[4]Figure 2 displays the *ex ante* returns to education under the alternative identification strategies.

Variable	Panel 1	Panel 2	Panel 3	Panel 4
Constant (δ_0)	-0.016 (0.171)	0.006 (0.175)	-0.028 (0.164)	-0.024 (0.155)
Local average income				
Higher education graduates	-0.01 (0.007)	-0.013* (0.008)	-0.01 (0.008)	-0.014* (0.008)
Local rate of honours		-0.014 (0.031)		
Secondary schooling track				
L	-0.128*** (0.046)	-0.132***(0.049)	-0.117** (0.059)	-0.142*** (0.054)
ES	-0.154*** (0.05)	-0.162*** (0.052)	-0.15** (0.063)	-0.172*** (0.058)
S	-0.146*** (0.051)	-0.164*** (0.054)	-0.135** (0.066)	-0.175*** (0.061)
Vocational	0.227 (0.226)	0.351** (0.173)	0.251 (0.175)	0.293* (0.165)
Technical	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Born abroad	-0.02 (0.02)	-0.032 (0.02)	-0.03 (0.022)	-0.032 (0.021)
Father born abroad	0 (0.01)	-0.005 (0.011)	-0.006 (0.012)	-0.005 (0.011)
Mother born abroad	-0.011 (0.011)	-0.006 (0.012)	-0.009 (0.014)	-0.009 (0.013)
Entering the labor market in 1998 (relative to 1992)	-0.094*** (0.034)	-0.106** (0.045)	-0.113** (0.055)	-0.12** (0.051)
Male	-0.061*** (0.012)	-0.043*** (0.008)	-0.044*** (0.009)	-0.038*** (0.009)
Father's profession				
Farmer	-0.02 (0.016)	-0.022 (0.016)	-0.018 (0.018)	-0.023 (0.017)
Tradesman	-0.021** (0.009)	-0.026** (0.013)	-0.02* (0.012)	-0.025** (0.011)
Executive	-0.051** (0.023)	-0.053** (0.022)	-0.043* (0.024)	-0.055** (0.022)
Intermediate occupation	-0.034** (0.014)	-0.04*** (0.015)	-0.03** (0.013)	-0.035*** (0.012)
Blue collar	-0.009 (0.007)	-0.007 (0.007)	-0.005 (0.009)	-0.004 (0.008)
Other	-0.016 (0.011)	-0.018 (0.011)	-0.021* (0.012)	-0.023** (0.011)
White collar	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Mother's profession				
Farmer	0.049 (0.034)	0.045 (0.03)	0.049 (0.039)	0.057 (0.037)
Tradesman	-0.008 (0.01)	0.002 (0.012)	-0.006 (0.012)	-0.003 (0.011)
Executive	-0.017 (0.012)	-0.018 (0.012)	-0.017 (0.015)	-0.023* (0.014)
Intermediate occupation	-0.018 (0.012)	-0.017 (0.011)	-0.019 (0.012)	-0.019* (0.011)
Blue collar	0.011 (0.007)	0.017* (0.01)	0.016 (0.01)	0.019* (0.01)
Other	-0.008 (0.007)	-0.01 (0.006)	-0.009 (0.007)	-0.01 (0.007)
White collar	*Ref.*	*Ref.*	*Ref.*	*Ref.*
Age in 6th grade				
≤ 10	-0.037* (0.021)	-0.039** (0.019)	-0.033 (0.025)	-0.047** (0.024)
11	*Ref.*	*Ref.*	*Ref.*	*Ref.*
≥ 12	0.06* (0.036)	0.05** (0.024)	0.048* (0.028)	0.056** (0.026)
Paris region	-0.024* (0.012)	-0.023* (0.012)	-0.018 (0.014)	-0.03** (0.012)
Vocational $\times$ ...				
Entering the labor market in 1998	0.019 (0.023)	0.04* (0.023)	0.02 (0.026)	0.034 (0.024)
Male	0.019 (0.016)	0.004 (0.015)	0.008 (0.015)	0.003 (0.014)
Paris region	0.045* (0.025)	0.052* (0.028)	0.038 (0.032)	0.059** (0.029)

In Panel 1, the higher education dropouts are excluded from the sample. In Panel 2, the local rate of honours is included in the estimation. In Panel 3 and 4, the streams of income were computed using ($\tau = 0.97, B = 25$) and ($\tau = 0.95, B = 30$) respectively. Standard errors, presented in parentheses, were computed by bootstrap with 200 sample replicates. Significance levels: *** (1%), ** (5%) and * (10%).

Table 4: Estimates of non-pecuniary factors: robustness checks.

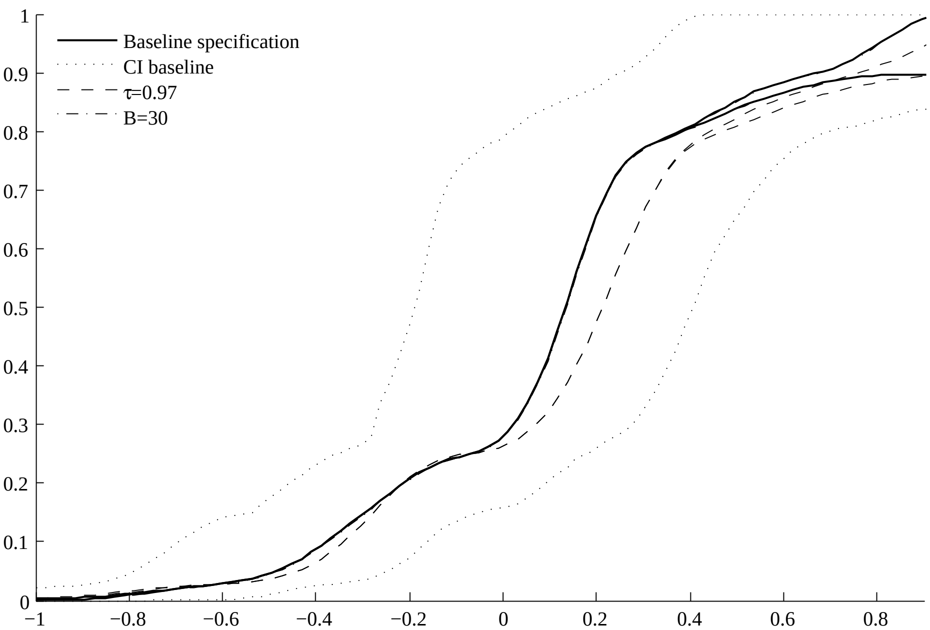

Figure 1: *Ex ante* returns to higher education under alternative computations of the streams of earnings.

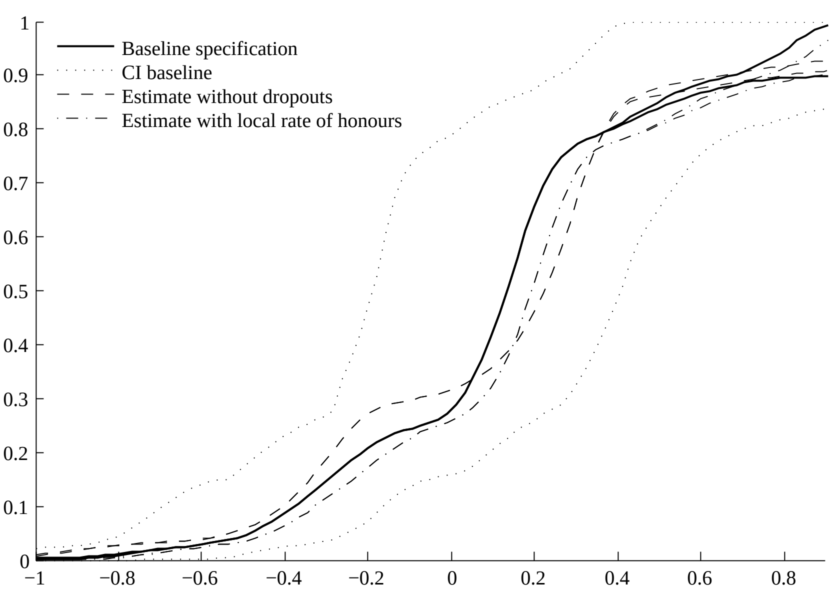

Figure 2: *Ex ante* returns to higher education: robustness of the instrumental strategy.

3 Proofs

3.1 Proofs of the main results

Theorem 2.1

First, by Assumption 2.2, $T(., x_{-1})$ is identified everywhere on the support of X_1 conditional on $X_{-1} = x_{-1}$, for almost all x_{-1}. Next, by Assumption 2.3 and the fundamental theorem of calculus,

$$\frac{\partial q_0}{\partial x_1}(x_1, x_{-1}) = -\frac{\partial(T+G)}{\partial x_1}(x_1, x_{-1}) f_{\eta_\Delta}\left(T(x_1, x_{-1}) + G(x_1, x_{-1})\right), \tag{3.1}$$

for almost all x_{-1} in the support of X_{-1} and all x_1 in the support of X_1 conditional on $X_{-1} = x_{-1}$. Fix x_{-1} so that Equations (2.6) in the main text and (3.1) above hold (the set of such x_{-1} being of probability one). By Assumption 2.3, $\frac{\partial q_0}{\partial x_1}(x_1, x_{-1}) \neq 0$ as soon as $\frac{\partial(T+G)}{\partial x_1}(x_1, x_{-1}) \neq 0$. Hence, by Equation (2.6) in the main text, $G(., x_{-1})$ is identified everywhere on the set

$$\mathcal{A}_{x_{-1}} = \{x_1 : \frac{\partial(T+G)}{\partial x_1}(x_1, x_{-1}) \neq 0\}.$$

By continuity of $G(., x_{-1})$, it is identified on the closure of $\mathcal{A}_{x_{-1}}$. If this set is equal to the support of X_1 conditional on $X_{-1} = x_{-1}$, then G is identified. Otherwise, let us consider $x_1 \notin \mathcal{A}_{x_{-1}}$. Because $\mathcal{A}_{x_{-1}} \neq \emptyset$ by Assumption 2.4, either $\mathcal{A}_{x_{-1}} \cap (-\infty, x_1)$ or $\mathcal{A}_{x_{-1}} \cap (x_1, \infty)$ is nonempty. Suppose without loss of generality that $\mathcal{A}_{x_{-1}} \cap (-\infty, x_1)$ is nonempty, and let $\overline{x}_1$ denote its supremum. If $x_1 = \overline{x}_1$, then G is identified at x_1 because it is identified on the closure of $\mathcal{A}_{x_{-1}}$ which contains $\overline{x}_1$. Otherwise, we have by definition $\frac{\partial(T+G)}{\partial x_1}(., x_{-1}) = 0$ on $(\overline{x}_1, x_1]$. Thus, $[T+G](., x_{-1})$ is constant on $(\overline{x}_1, x_1]$ and, by continuity, on $[\overline{x}_1, x_1]$. Hence,

$$G(x_1, x_{-1}) = -T(x_1, x_{-1}) + G(\overline{x}_1, x_{-1}) + T(\overline{x}_1, x_{-1}).$$

The result follows because $G(\overline{x}_1, x_{-1})$, $T(x_1, x_{-1})$ and $T(\overline{x}_1, x_{-1})$ are identified.

Proposition 2.2

Recall that $\varepsilon_k = \eta_k + \nu_k$ for $k \in \{0, 1\}$. Because $E(\nu_k | X, \eta_0, \eta_1) = 0$, we have $E(\nu_k | X, D = k) = 0$. Moreover, by by Assumption 2.3, S_{η_Δ} is strictly decreasing.

Thus, by Assumptions 2.1 and 2.3,

$$\begin{aligned}
E(\varepsilon_1|D=1,X=x) &= \frac{E(\eta_1 D|X=x)}{P(D=1|X=x)} \\
&= \frac{E\left(\eta_1 \mathbb{1}\{\eta_\Delta \geq \psi_0(x)-\psi_1(x)+G(x)\}\right)}{P(D=1|X=x)} \\
&= \frac{E(\eta_1 \mathbb{1}\{S_{\eta_\Delta}(\eta_\Delta) \leq P(D=1|X=x)\})}{P(D=1|X=x)}.
\end{aligned}$$

In other terms, there exists a measurable function h such that $E(\varepsilon_1|D=1,X) = h(P(D=1|X))$. Now, by Assumption 2.6,

$$E(Y|D=1,X) = \psi_1(\widetilde{X}_1) + h(P(D=1|X)).$$

Suppose that there exists $\widetilde{\psi_1}$ and $\widetilde{h}$ such that

$$E(Y|D=1,X) = \widetilde{\psi_1}(\widetilde{X}_1) + \widetilde{h}(P(D=1|X)).$$

Then

$$(\widetilde{\psi_1} - \psi_1)(\widetilde{X}_1) + (\widetilde{h} - h)(P(D=1|X)) = 0$$

By the measurably separation condition, this implies that $\widetilde{\psi_1}$ and ψ_1 are almost surely equal up to a constant. This constant is identified by Assumption 2.5. Thus, ψ_1 is identified. ψ_0 can be recovered by the same argument.

Proposition 2.3

The proof relies on Theorem 2.1 of D'Haultfoeuille & Maurel (2012). Their Assumptions 1 and 2 are satisfied by Conditions (i) and (ii) of Assumption 2.7. All we have to check is that their Assumption 3 also holds. For that purpose, remark that for $k \in \{0,1\}$,

$$\begin{aligned}
P(D=k|X=x,Y_k=y) &= P(D=k|X=x,\varepsilon_k = y-\psi_k(x)) \\
&= P(\eta_k - \eta_{1-k} > \psi_{1-k}(x) - \psi_k(x) + G(x)|\eta_k + \nu_k = y - \psi_k(x)).
\end{aligned}$$

Thus, by Condition (iii) of Assumption 2.7,

$$\lim_{y\to\infty} P(D=k|X=x,Y_k=y) = 1, \text{ for all } x.$$

This implies Assumption 3 of D'Haultfoeuille & Maurel (2012), and the result follows.

Theorem 3.1

Before establishing the result, let us introduce some notations. Let $f(.,\zeta)$ denote the density of $X'\zeta$, $q(u,\zeta) = E(D|X'\zeta = u)$, $r(.,\zeta) = q(.,\zeta) \times f(.,\zeta)$ and define $f_0(.) = f(.,\zeta_0)$, $q_0(.) = q(.,\zeta_0)$ and $r_0(.) = q_0(.)f_0(.)$. Consider the kernel estimators

$$\widehat{f}(u,\zeta) = \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{u - X_i'\zeta}{h_n}\right)$$

and $\widehat{r}(.,\zeta) = \widehat{q}(.,\zeta) \times \widehat{f}(.,\zeta)$, where $\widehat{q}(.,\zeta)$ is defined by Equation (3.6) in the main text. Let us also define $Z_i(\zeta) = \mathbb{1}\{X_i \in \mathcal{X}\}h(X_i'\zeta)$ and, for any $\mu = (r(.), f(.), \zeta, \widetilde{\beta}_0, \widetilde{\beta}_1)$,

$$V_i(\mu) = D_i X_i'\zeta - \int_{u_0}^{X_i'\zeta} \frac{r(u)}{f(u)} du.$$

We then let $W_i(\mu) = (1, D_i, V_i(\mu))'$. Thus, $\widehat{W}_i = W_i(\widehat{\mu})$ and $W_i = W_i(\mu_0)$, with $\widehat{\mu} = (\widehat{r}(.,\widehat{\zeta}), \widehat{f}(.,\widehat{\zeta}), \widehat{\zeta}, \widehat{\beta}_0, \widehat{\beta}_1)$ and $\mu_0 = (r_0, f_0, \zeta_0, \beta_0, \beta_1)$. Similarly, let

$$\varepsilon_i(\mu) = Y_i - X_i'\left(D_i\widetilde{\beta}_1 + (1 - D_i)\widetilde{\beta}_0\right).$$

Eventually, let $g(A_i, \theta, \mu) = Z_i(\zeta)(\varepsilon_i(\mu) - W_i(\mu)'\theta)$ and $g(A_i, \mu) = g(A_i, \theta_0, \mu)$, with $A_i = (D_i, Y_i, X_i)$. Then $E[g(A, \mu_0)] = 0$ and

$$\sum_{i=1}^n g(A_i, \widehat{\theta}, \widehat{\mu}) = 0.$$

Thus, $\widehat{\theta}$ is a two step GMM estimator with a nonparametric first step estimator, and we follow Newey & McFadden (1994)'s outline for establishing asymptotic normality. Some differences arise however because of the estimation of ζ in the nonparametric estimator of q_0. The proof of the theorem proceeds in three steps.

Step 1. We first show that $\mu \mapsto \sum_{i=1}^n g(A_i, \mu)$ can be linearized in a convenient way. Recalling that $U_i = X_i'\zeta_0$, we let

$$\begin{aligned} G(A_i, \mu) \quad = \quad & \xi_i \frac{\partial Z_i}{\partial \zeta}(\zeta_0)'\zeta + Z_i(\zeta_0)\Bigg[- X_i'(D_i\widetilde{\beta}_1 + (1 - D_i)\widetilde{\beta}_0) - \Bigg(D_i X_i'\zeta \\ & - q_0(U_i)X_i'\zeta - \int_{u_0}^{U_i} \frac{\partial q}{\partial \zeta}(u,\zeta_0)'\zeta + \frac{1}{f_0(u)}\left(r(u) - q_0(u)f(u)\right) du\Bigg)\alpha_0\Bigg]. \end{aligned}$$

Note that $\partial q/\partial\zeta(.,\zeta_0)$ exists under Assumptions 2.3 and 3.2, by Lemma 3.1 below. Let us also define $\widetilde{\mu} = (\widetilde{r}, \widetilde{f}, \widehat{\zeta}, \widehat{\beta}_0, \widehat{\beta}_1)$ where $\widetilde{r} = \widehat{r}(.,\zeta_0)$ and $\widetilde{f} = \widehat{f}(.,\zeta_0)$. We shall prove that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[g(A_i, \widehat{\mu}) - g(A_i, \mu_0) - G(A_i, \widetilde{\mu} - \mu_0)\right] = o_P(1). \qquad (3.2)$$

For that purpose, we use the decomposition

$$g(A_i,\widehat{\mu}) - g(A_i,\mu_0) - G(A_i,\widetilde{\mu}-\mu_0) = R_{1i} + R_{2i} + R_{3i} + R_{4i} + R_{5i}$$

where, denoting by $h'(.)$ the vector of derivatives of $h(.)$ and $\widetilde{q} = \widetilde{r}/\widetilde{f}$, we let

$$\begin{aligned}
R_{1i} &= \xi_i \mathbb{1}\{X_i \in \mathcal{X}\}\left(h(\widehat{U}_i) - h(U_i) - (\widehat{U_i} - U_i)h'(U_i)\right),\\
R_{2i} &= \alpha_0 Z_i(\zeta_0)\left[\int_{U_i}^{\widehat{U}_i} \widehat{q}(u,\widehat{\zeta})du - q_0(U_i)(\widehat{U}_i - U_i)\right],\\
R_{3i} &= \alpha_0 Z_i(\zeta_0)\int_{u_0}^{U_i} \widehat{q}(u,\widehat{\zeta}) - \widetilde{q}(u) - \frac{\partial q}{\partial \zeta}(u,\zeta_0)'(\widehat{\zeta}-\zeta_0)du,\\
R_{4i} &= \alpha_0 Z_i(\zeta_0)\int_{u_0}^{U_i} \widetilde{q}(u) - q_0(u) - \frac{1}{f_0(u)}\left(\widetilde{r}(u) - r_0(u) - q_0(u)(\widetilde{f}(u) - f_0(u))\right)du,\\
R_{5i} &= [\varepsilon_i(\widehat{\mu}) - \varepsilon_i(\mu_0) - (W_i(\widehat{\mu}) - W_i(\mu_0))'\theta_0]\left[Z_i(\widehat{\zeta}) - Z_i(\zeta_0)\right].
\end{aligned}$$

We now check that for all $k \in \{1,...,5\}$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ki} = o_P(1)$.

$-$ R_{1i}: by Assumption 3.2, there exists C_0 such that $\|X\| \leq C_0$, where $\|.\|$ denotes the euclidian norm. Then, by the Cauchy-Schwarz inequality, $|\widehat{U_i} - U_i| \leq C_0||\widehat{\zeta} - \zeta_0||$. Thus, by Assumptions 3.4 and 3.7,

$$\begin{aligned}
\sqrt{n}\max_{i=1,...,n}\left|h(\widehat{U_i}) - h(U_i) - (\widehat{U}_i - U_i)h'(U_i)\right| &\leq \sqrt{n}M\max_{i=1,...,n}|\widehat{U}_i - U_i|^2\\
&\leq MC_0^2\sqrt{n}||\widehat{\zeta} - \zeta_0||^2\\
&= o_P(1),
\end{aligned}$$

where $M = \|\max|h''|\|$. Besides, $\sum_{i=1}^n |\xi_i|/n = O_P(1)$. Thus,

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{1i}\right\| = o_P(1).$$

$-$ R_{2i}: Let $\mathcal{S}_0 = \{x'\zeta_0, x \in \mathcal{X}\}$. By definition, $\mathcal{S}_0 \subsetneq \mathcal{S}$, where $\mathcal{S}$ denotes the support of U. Besides, by definition, $Z_i(\zeta_0) = Z_i(\zeta_0)\mathbb{1}\{U_i \in \mathcal{S}_0\}$. Moreover, for all i such that $\widehat{U}_i \in \mathcal{S}_0$, there exists, by the mean value theorem, $\widetilde{U}_i = tU_i + (1-t)\widehat{U}_i$, with $t \in [0,1]$, such that $\int_{U_i}^{\widehat{U}_i} q_0(u)du = q_0(\widetilde{U}_i)(\widehat{U}_i - U_i)$. Thus, when $\widehat{U}_i \in \mathcal{S}_0$,

$$\begin{aligned}
\|R_{2i}\| &= \left\|\alpha_0 Z_i(\zeta_0)\mathbb{1}\{U_i \in \mathcal{S}_0\}\left\{\int_{U_i}^{\widehat{U}_i}\left[\widehat{q}(u,\widehat{\zeta}) - q_0(u)\right]du + \int_{U_i}^{\widehat{U}_i} q_0(u)du - q_0(U_i)(\widehat{U}_i - U_i)\right\}\right\|\\
&\leq C_1\left|\widehat{U}_i - U_i\right|\left[\sup_{u\in\mathcal{S}_0}\left|\widehat{q}(u,\widehat{\zeta}) - q_0(u)\right| + \max_{i:\widehat{U}_i\in\mathcal{S}}\left|q_0(\widetilde{U}_i) - q_0(U_i)\right|\right]\\
&\leq C_0C_1\left\|\widehat{\zeta} - \zeta_0\right\|\left[\sup_{u\in\mathcal{S}_0}\left|\widehat{q}(u,\widehat{\zeta}) - q_0(u)\right| + \max_{i:\widehat{U}_i\in\mathcal{S}}\left|q_0(\widetilde{U}_i) - q_0(U_i)\right|\right],
\end{aligned}$$

where $C_1 > 0$ is a constant such that $\|\alpha_0 Z_i(\zeta_0)\| \leq C_1$, which exists by Assumptions 3.2 and 3.7. Besides, because $\widehat{q}(.,\widehat{\zeta})$ and $q_0(.)$ are bounded above by 1, we have, when $\widehat{U}_i \notin \mathcal{S}_0$,

$$\|R_{2i}\| \leq 2C_0C_1 \left\|\widehat{\zeta} - \zeta_0\right\| \mathbb{1}\{U_i \in \mathcal{S}_0\}.$$

Hence,

$$\begin{aligned}\|R_{2i}\| \quad \leq \quad & C_0C_1 \left\|\widehat{\zeta} - \zeta_0\right\| \left[\sup_{u\in\mathcal{S}_0} \left|\widehat{q}(u,\widehat{\zeta}) - q_0(u)\right| + \max_{i:\widehat{U}_i\in\mathcal{S}} \left|q_0(\widetilde{U}_i) - q_0(U_i)\right|\right. \\ & \left. +2\mathbb{1}\{U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\}\right]. \end{aligned} \tag{3.3}$$

By Assumption 3.4, $\sqrt{n}\left\|\widehat{\zeta} - \zeta_0\right\| = O_P(1)$. Let us now show that the term into brackets in (3.3) is a $o_P(1)$. By Lemma 3.2 below, $\sup_{u\in\mathcal{S}_0} |\widehat{q}(u,\widehat{\zeta}) - q_0(u)| = o_P(1)$. Now fix $\varepsilon > 0$. Because $q_0(.)$ is continuous by Assumption 2.3 and $\mathcal{S}$ is compact, $q_0(.)$ is uniformly continuous on $\mathcal{S}$. Thus, there exists $\delta > 0$ such that for all $(u,v) \in \mathcal{S}^2$ satisfying $|u - v| \leq \delta$, we have $|q_0(u) - q_0(v)| \leq \varepsilon$. As a consequence,

$$P\left(\max_{i:\widehat{U}_i\in\mathcal{S}} \left|q_0(\widetilde{U}_i) - q_0(U_i)\right| \leq \varepsilon\right) \geq P\left(\max_{i:\widehat{U}_i\in\mathcal{S}} \left|\widetilde{U}_i - U_i\right| \leq \delta\right).$$

Because $|\widetilde{U}_i - U_i| \leq |\widehat{U}_i - U_i| \leq C_0\left\|\widehat{\zeta} - \zeta_0\right\|$, the right-hand side tends to one. This proves that

$$\max_{i:\widehat{U}_i\in\mathcal{S}} \left|q_0(\widetilde{U}_i) - q_0(U_i)\right| = o_P(1).$$

It remains to show that

$$\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\} = o_P(1). \tag{3.4}$$

For all $\delta > 0$, let $\mathcal{S}_\delta = \{s \in \mathcal{S}_0 / \exists s' \notin \mathcal{S}_0 / |s - s'| < \delta\}$. Fix $\varepsilon > 0$ and let $K > 0$ be such that $P(U_i \in \mathcal{S}_K) < \varepsilon/2$. For n large enough, $P(C_0\left\|\widehat{\zeta} - \zeta_0\right\| > K) < \varepsilon/2$. Because $|U_i - \widehat{U}_i| \leq C_0\left\|\widehat{\zeta} - \zeta_0\right\|$, we have, for n large enough,

$$\begin{aligned}P\left(U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\right) \quad & \leq \quad \frac{\varepsilon}{2} + P\left(U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0, C_0||\widehat{\zeta} - \zeta_0|| \leq K\right) \\ & \leq \quad \frac{\varepsilon}{2} + P(U_i \in \mathcal{S}_K) \\ & \leq \quad \varepsilon.\end{aligned}$$

Because ε was arbitrary, this proves that

$$E\left[\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{U_i \in \mathcal{S}_0, \widehat{U}_i \notin \mathcal{S}_0\}\right|\right] \to 0.$$

This establishes (3.4) since convergence in L^1 implies convergence in probability. As a result, $\sum_{i=1}^n R_{2i}/\sqrt{n} = o_P(1)$.

$-\, R_{3i}$: By the mean value theorem, there exists $\widetilde{\zeta}_u$ in the segment between ζ_0 and $\widehat{\zeta}$ such that

$$\widehat{q}(u,\widehat{\zeta}) - \widetilde{q}(u) = \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u)'(\widehat{\zeta}-\zeta_0).$$

Because U_i is bounded, there exists C_2 such that $|U_i - u_0| < C_2$. Thus,

$$\begin{aligned} |R_{3i}| &= \|\alpha_0 Z_i(\zeta_0)\| \left| \left[\int_{u_0}^{U_i} \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u) - \frac{\partial q}{\partial \zeta}(u,\zeta_0) du \right]' (\widehat{\zeta}-\zeta_0) \right| \mathbb{1}\{U_i \in \mathcal{S}_0\} \\ &\leq C_1 C_2 \left\| \widehat{\zeta} - \zeta_0 \right\| \sup_{u \in \mathcal{S}_0} \left\| \frac{\partial \widehat{q}}{\partial \zeta}(u,\widetilde{\zeta}_u) - \frac{\partial q}{\partial \zeta}(u,\zeta_0) \right\|. \end{aligned}$$

The supremum tends to zero in probability by Lemma 3.2. As a result, $\sum_{i=1}^n R_{3i}/\sqrt{n} = o_P(1)$.

$-\, R_{4i}$: following Newey & McFadden (1994, p. 2204), we have

$$\begin{aligned} |R_{4i}| &\leq C_1 \mathbb{1}\{U_i \in \mathcal{S}_0\} \int_{u_0}^{U_i} \frac{1}{\widetilde{f}(u) f_0(u)} [1 + |q_0(u)|] \left[|\widetilde{f}(u) - f_0(u)|^2 + |\widetilde{r}(u) - r_0(u)|^2 \right] du \\ &\leq \frac{2C_1C_2}{\inf_{u\in\mathcal{S}_0} \widetilde{f}(u) \inf_{u\in\mathcal{S}_0} f_0(u)} \left[\left(\sup_{u\in\mathcal{S}_0} |\widetilde{f}(u) - f_0(u)| \right)^2 + \left(\sup_{u\in\mathcal{S}_0} |\widetilde{r}(u) - r_0(u)| \right)^2 \right] \quad (3.5) \end{aligned}$$

Assumption 3.2 implies that the density of U_i is positive in the interior of $\mathcal{S}$. Thus, $\inf_{u\in\mathcal{S}_0} f_0(u) > 0$. By uniform consistency of $\widetilde{f}$ on $\mathcal{S}_0$ (see, e.g., Lemma 8.10 of Newey & McFadden, 1994) the ratio in the right-hand side of (3.5) is a $O_P(1)$. Thus it suffices to show that $\sup_{u\in\mathcal{S}_0} |\widetilde{f}(u) - f_0(u)| = o_P(n^{-1/4})$ and similarly for $\widetilde{r}$. The result follows from Assumption 3.6, the rate condition on h_n and Lemma 8.10 of Newey & McFadden (1994).

$-\, R_{5i}$: first, note that

$$\begin{aligned} & \left| \varepsilon_i(\widehat{\mu}) - \varepsilon_i(\mu_0) - (W_i(\widehat{\mu}) - W_i(\mu_0))' \theta_0 \right| \mathbb{1}\{X_i \in \mathcal{X}\} \\ = & \left| X_i'(D_i(\beta_1 - \widehat{\beta}_1) + (1-D_i)(\beta_0 - \widehat{\beta}_0)) + \left(D_i(U_i - \widehat{U}_i) + \int_{U_i}^{\widehat{U}_i} \widehat{q}(u,\widehat{\zeta}) du \right. \right. \\ & \left. \left. + \int_{u_0}^{U_i} \left[\widehat{q}(u,\widehat{\zeta}) - q_0(u) \right] du \right) \alpha_0 \right| \mathbb{1}\{X_i \in \mathcal{X}\} \\ \leq & \ C_0 \left(\left\| \widehat{\beta}_1 - \beta_1 \right\| + \left\| \widehat{\beta}_0 - \beta_0 \right\| + 2|\alpha_0| \left\| \widehat{\zeta} - \zeta_0 \right\| \right) + C_2 |\alpha_0| \sup_{u\in\mathcal{S}_0} |\widehat{q}(u,\widehat{\zeta}) - q_0(u)|. \end{aligned}$$

where the first term of the upper bound follows from the Cauchy-Schwarz inequality.

Besides, with probability approaching one, there exists a compact which contains $\widehat{U}_i$ and U_i for all i. Thus, because h' is continuous, there exists $C_3 > 0$ such that, with probability approaching one,

$$\left\| Z_i(\widehat{\zeta}) - Z_i(\zeta_0) \right\| \leq C_3 \left\| \widehat{\zeta} - \zeta_0 \right\|.$$

Hence, with probability approaching one,

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} R_{5i}\right| \leq \left[C_0 C_3 \sqrt{n}\left\|\widehat{\zeta}-\zeta_0\right\|\right]\left[\left\|\widehat{\beta}_1-\beta_1\right\|+\left\|\widehat{\beta}_0-\beta_0\right\|+2|\alpha_0|\left\|\widehat{\zeta}-\zeta_0\right\|\right.$$
$$\left.+C_2|\alpha_0|\sup_{u\in\mathcal{S}_0}|\widehat{q}(u,\widehat{\zeta})-q_0(u)|\right].$$

By Assumption 3.4, the first term into brackets in the right-hand side is a $O_P(1)$. By Lemma 3.2 and Assumptions 3.4 and 3.5, the second term is a $o_P(1)$. The result follows.

Step 2. Now, let us show that $1/\sqrt{n}\sum_{i=1}^{n} G(A_i,\widetilde{\mu}-\mu_0)$ can be linearized. Let $\kappa_0 = (\zeta_0,\beta_1,\beta_0)'$ and $\widehat{\kappa} = (\widehat{\zeta},\widehat{\beta_1},\widehat{\beta_0})'$. We have

$$G(A_i,\widetilde{\mu}-\mu_0) = P_i'\left(\widehat{\kappa}-\kappa_0\right)+\widetilde{G}(A_i,\widetilde{r},\widetilde{f}),$$

with $P_i = (P_{1i},P_{2i},P_{3i})'$ and

$$\begin{aligned}
P_{1i} &= \xi_i\frac{\partial Z_i}{\partial\zeta}(\zeta_0)'-\alpha_0 Z_i(\zeta_0)\left(D_iX_i'-q_0(U_i)X_i'-\int_{u_0}^{U_i}\frac{\partial q}{\partial\zeta'}(u,\zeta_0)du\right)\\
P_{2i} &= -Z_i(\zeta_0)D_iX_i'\\
P_{3i} &= -Z_i(\zeta_0)(1-D_i)X_i'\\
\widetilde{G}(A_i,\widetilde{r},\widetilde{f}) &= \alpha_0 Z_i(\zeta_0)\int_{u_0}^{U_i}(1/f_0(u))(\widetilde{r}(u)-q_0(u)\widetilde{f}(u))du.
\end{aligned}$$

By the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} P_i \xrightarrow{P} E\left[P\right].$$

Moreover, by Assumptions 3.4 and 3.5,

$$\sqrt{n}\left(\widehat{\kappa}-\kappa_0\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\chi_i,\chi_{1i},\chi_{0i}\right)'+o_P(1).$$

Thus,

$$\left(\frac{1}{n}\sum_{i=1}^{n} P_i\right)'\sqrt{n}(\widehat{\kappa}-\kappa_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_{1i}+o_P(1), \tag{3.6}$$

where

$$\Omega_{1i} = E[P]'\left(\chi_i,\chi_{1i},\chi_{0i}\right)'. \tag{3.7}$$

Thus, it suffices to focus on the nonparametric part of G, $\widetilde{G}(A_i,\widetilde{r},\widetilde{f})$. The main insight here is that $\widetilde{G}$ is nearly the linearized part of the consumer surplus example of Newey & McFadden (1994, p. 2204), except that their b is replaced by U_i. Thus, it suffices to modify slightly their proof (see Newey & McFadden, 1994, p. 2211) to satisfy Conditions (ii), (iii) and (iv) as well as the technical requirements of their Theorem 8.11. As a result, we get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{G}(A_i,r,f)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Omega_{2i}+o_P(1), \tag{3.8}$$

where $\Omega_{2i}=\alpha_0 Z_i(\zeta_0)(1-F_0(U_i))\mathbb{1}\{U_i\geq u_0\}(D_i-q_0(U_i))/f_0(U_i)$, $F_0(.)$ denoting the cumulative distribution function of U. The result follows.

Step 3. Eventually, we establish the asymptotic normality of $\widehat{\theta}$. By (3.2), (3.6) and (3.8) and the central limit theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g(A_i,\widehat{\mu})\xrightarrow{d}\mathcal{N}\left(0,\,V(g(A,\mu_0)+\Omega_{11}+\Omega_{21})\right).$$

Thus, by definition of $\widehat{\theta}$ and $g(A_i,\theta,\widehat{\mu})$,

$$\left[\frac{1}{n}\sum_{i=1}^{n}Z_i(\widehat{\zeta})W_i(\widehat{\mu})'\right]\sqrt{n}(\widehat{\theta}-\theta_0)\xrightarrow{d}\mathcal{N}\left(0,\,V(g(A,\mu_0)+\Omega_{11}+\Omega_{21})\right).$$

Now,

$$Z_i(\widehat{\zeta})W_i(\widehat{\mu})'=Z_i(\zeta_0)W_i(\mu_0)'+Z_i(\widehat{\zeta})(W_i(\widehat{\mu})-W_i(\mu_0))'+(Z_i(\widehat{\zeta})-Z_i(\zeta_0))W_i(\mu_0)'.$$

Besides, by Assumption 3.7, $\left\|Z_i(\widehat{\zeta})-Z_i(\zeta_0)\right\|\leq C_3\left\|\widehat{\zeta}-\zeta_0\right\|$ for a given $C_3>0$. Moreover, reasoning as with R_{5i}, we get

$$\|W_i(\widehat{\mu})-W_i(\mu_0)\|\leq 2C_0\left\|\widehat{\zeta}-\zeta_0\right\|+C_2\sup_{u\in\mathcal{S}_0}|\widehat{q}(u,\widehat{\zeta})-q_0(u)|.$$

Finally, $\|W_i(\mu_0)\|$ and $\left\|Z_i(\widehat{\zeta})\right\|$ are bounded with probability approaching one. As a result,

$$\frac{1}{n}\sum_{i=1}^{n}Z_i(\widehat{\zeta})W_i(\widehat{\mu})'=\frac{1}{n}\sum_{i=1}^{n}Z_i(\zeta_0)W_i(\mu_0)'+o_P(1).$$

Thus, by the weak law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}Z_i(\widehat{\zeta})W_i(\widehat{\mu})'\xrightarrow{P}E(Z(\zeta_0)W(\mu_0)')=E(ZW').$$

Eventually, by Slutski's lemma, and given that $g(A,\mu_0)=Z\xi$,

$$\sqrt{n}(\widehat{\theta}-\theta_0)\xrightarrow{d}\mathcal{N}\left(0,\,E(ZW')^{-1}V(Z\xi+\Omega_{11}+\Omega_{21})E(WZ')^{-1}\right).$$

This concludes the proof.

3.2 Technical lemmas

Lemma 3.1 *Suppose that Assumptions 2.3 and 3.2 hold. Then, for all $u \in \mathcal{S}$, the support of U, $\zeta \mapsto f(u,\zeta)$ and $\zeta \mapsto r(u,\zeta)$, the density of $X'\zeta$ and the derivative of $u \mapsto E(D\mathbb{1}\{X'\zeta \leq u\})$ respectively, admit partial derivatives at ζ_0 which satisfy:*

$$\frac{\partial f}{\partial \zeta}(u,\zeta_0) = -\left(E\left[X|U=u\right] f_0(u)\right)' \tag{3.9}$$

$$\frac{\partial r}{\partial \zeta}(u,\zeta_0) = -\left(E\left[DX|U=u\right] f_0(u)\right)' \tag{3.10}$$

Proof: let $X_{-m} = (X_1, ..., X_{m-1}, X_{m+1}..., X_p)$ and $f_{X_m|X_{-m}}(.,x)$ (resp. $F_{X_m|X_{-m}}(.,x)$) denote the density (resp. cumulative distribution function) of X_m conditional on $X_{-m} = x$. Let also δ_k denote the vector of dimension p, with 1 at the k-th component and 0 elsewhere. We have

$$f(u,\zeta+t\delta_k) = \left| \begin{array}{ll} E\left[f_{X_m|X_{-m}}\left(\frac{u-X'_{-m}\zeta_{-m}-tX_k}{\zeta_m}, X_{-m}\right)\right] & \text{if } k \neq m, \\ E\left[f_{X_m|X_{-m}}\left(\frac{u-X'_{-m}\zeta_{-m}}{\zeta_m+t}, X_{-m}\right)\right] & \text{if } k = m. \end{array}\right.$$

Thus, by Assumption 3.2 and dominated convergence, $\zeta \mapsto f(u,\zeta)$ admits continuous partial derivatives. Now, let $F(.,\zeta)$ denote the cumulative distribution function of $X'\zeta$. We have,

$$F(u,\zeta+t\delta_k) = \left| \begin{array}{ll} E\left[F_{X_m|X_{-m}}\left(\frac{u-X'_{-m}\zeta_{-m}-tX_k}{\zeta_m}, X_{-m}\right)\right] & \text{if } k \neq m, \\ E\left[F_{X_m|X_{-m}}\left(\frac{u-X'_{-m}\zeta_{-m}}{\zeta_m+t}, X_{-m}\right)\right] & \text{if } k = m. \end{array}\right.$$

Thus, by Assumption 3.2 and dominated convergence, $\zeta \mapsto F(u,\zeta)$ admits continuous partial derivatives, and after some rearrangements,

$$\frac{\partial F}{\partial \zeta_k}(u,\zeta_0) = -E\left[X_k|U=u\right] f_0(u).$$

By Assumption 3.2 once more, $u \mapsto \partial F/\partial \zeta_k(u,\zeta_0)$ is continuously differentiable and

$$\frac{\partial^2 F}{\partial u \partial \zeta}(u,\zeta_0) = -\left(E\left[X|U=u\right] f_0(u)\right)'.$$

Then (3.9) follows from $\partial f/\partial \zeta = \partial^2 F/\partial \zeta \partial u = \partial^2 F/\partial u \partial \zeta$.

The proof of (3.10) is similar, except that we use $G_0(u,\zeta) = E(D\mathbb{1}\{X'\zeta \leq u\})$ instead of $F(u,\zeta)$. The partial derivatives of $\zeta \mapsto G_0(u,\zeta)$ exist and satisfy

$$\begin{aligned} \frac{\partial G_0}{\partial \zeta}(u,\zeta) &= -E\left(DX|U=u\right) f_0(u) \\ &= -S_{\eta_\Delta}(u+\delta_0) E\left(X|U=u\right) f_0(u). \end{aligned}$$

Then differentiability of $u \mapsto \partial G_0/\partial\zeta(u,\zeta)$ stems from Assumptions 2.3 and 3.2. Equation (3.10) follows from the same argument as previously.

Lemma 3.2 *Suppose that $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and Assumptions 3.2 and 3.6 hold. Then, for all closed interval $\mathcal{S}'$ strictly included in the interior of $\mathcal{S}$ and for all $\zeta_{u,n}$ such that $\sup_{u\in\mathcal{S}'}\|\zeta_{u,n}-\zeta_0\| = O_P(1/\sqrt{n})$, we have,*

$$\sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - q_0(u)| \;=\; o_P(1) \tag{3.11}$$

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_{u,n}) - \frac{\partial q}{\partial\zeta}(u,\zeta_0)\right\| \;=\; o_P(1) \tag{3.12}$$

Proof: we first write

$$\sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - q_0(u)| \le \sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - \widehat{q}(u,\zeta_0)| + \sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_0) - q_0(u)| \tag{3.13}$$

Let us first consider the the first term of the r.h.s. Since $|\widehat{q}(u,\zeta_{u,n})| \le 1$, we have

$$\begin{aligned}
\sup_{u\in\mathcal{S}'}|\widehat{q}(u,\zeta_{u,n}) - \widehat{q}(u,\zeta_0)| &= \sup_{u\in\mathcal{S}'}\frac{\left|(\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)) + \widehat{q}(u,\zeta_{u,n})(\widehat{f}(u,\zeta_0) - \widehat{f}(u,\zeta_{u,n}))\right|}{\widehat{f}(u,\zeta_0)} \\
&\le \sup_{u\in\mathcal{S}'}\frac{1}{\widehat{f}(u,\zeta_0)}\left[|\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)| + \left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right|\right] \\
&\le \frac{1}{\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0)}\left[\sup_{u\in\mathcal{S}'}|\widehat{r}(u,\zeta_{u,n}) - \widehat{r}(u,\zeta_0)|\right. \\
&\qquad \left. + \sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right|\right].
\end{aligned} \tag{3.14}$$

Let us prove that

$$\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| = o_P(1) \tag{3.15}$$

The proof for $\widehat{r}$ is similar. By Assumption 3.6, there exists $C_4 > 0$ such that $|K(u) - K(v)| \le C_4|u-v|$. Thus,

$$\begin{aligned}
\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| &\le \frac{1}{nh_n}\sum_{i=1}^n\left|K\left(\frac{u - X_i'\zeta_{u,n}}{h_n}\right) - K\left(\frac{u - X_i'\zeta_0}{h_n}\right)\right| \\
&\le \frac{C_4C_0\|\zeta_{u,n}-\zeta_0\|}{h_n^2} \\
&\le \frac{C_4C_0\sup_{u\in\mathcal{S}'}\|\zeta_{u,n}-\zeta_0\|}{h_n^2} = O_p\left(\frac{1}{\sqrt{n}h_n^2}\right).
\end{aligned}$$

This establishes (3.15) since $nh_n^4 \to \infty$. Because

$$\inf_{u\in\mathcal{S}'}\widehat{f}(u,\zeta_0) \ge -\sup_{u\in\mathcal{S}'}\left|\widehat{f}(u,\zeta_{u,n}) - \widehat{f}(u,\zeta_0)\right| + \inf_{u\in\mathcal{S}'}f_0(u),$$

and because $\inf_{u\in\mathcal{S}'} f_0(u) > 0$ by Assumption 3.2, we also have

$$\frac{1}{\inf_{u\in\mathcal{S}'} \widehat{f}(u,\zeta_0)} = O_p(1).$$

By (3.14), the first term of (3.13) tends to zero.

As for the second term, we can obtain the same decomposition as (3.14). Then Assumptions 3.2 and 3.6, and conditions on h_n ensure that we can apply Lemma 8.10 of Newey & McFadden (1994), yielding $\sup_{u\in\mathcal{S}'} |\widehat{f}(u,\zeta_0) - f_0(u)| = o_P(1)$ and similarly for $\widehat{r}(.,\zeta_0)$. This establishes (3.11).

Now, let us turn to (3.12). We use the same decomposition as (3.13). First, let us establish that

$$\sup_{u\in\mathcal{S}'} \left| \frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) - \frac{\partial q}{\partial\zeta}(u,\zeta_0) \right| = o_P(1) \qquad (3.16)$$

We have

$$\frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) = \frac{1}{\widehat{f}(u,\zeta_0)} \left[\frac{\partial\widehat{r}}{\partial\zeta}(u,\zeta_0) - \widehat{q}(u,\zeta_0)\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0) \right].$$

and similarly for $\partial q/\partial\zeta(u,\zeta_0)$. Thus,

$$\begin{aligned} & \frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) - \frac{\partial q}{\partial\zeta}(u,\zeta_0) \\ = \; & \frac{1}{\widehat{f}(u,\zeta_0)} \left\{ \left[\frac{\partial\widehat{r}}{\partial\zeta}(u,\zeta_0) - \frac{\partial r}{\partial\zeta}(u,\zeta_0) \right] - \frac{\partial r}{\partial\zeta}(u,\zeta_0) \left[\frac{\widehat{f}(u,\zeta_0) - f_0(u)}{f_0(u)} \right] \right\} \\ & - \frac{\widehat{q}(u,\zeta_0)}{\widehat{f}(u,\zeta_0)} \left[\left(\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0) - \frac{\partial f}{\partial\zeta}(u,\zeta_0) \right) - \frac{\partial f/\partial\zeta(u,\zeta_0)}{f_0(u)} \left(\widehat{f}(u,\zeta_0) - f_0(u) \right) \right] \\ & - \frac{\partial f/\partial\zeta(u,\zeta_0)}{f_0(u)} \left(\widehat{q}(u,\zeta_0) - q_0(u) \right). \end{aligned}$$

By what precedes, $\inf_{u\in\mathcal{S}'} \widehat{f}(u,\zeta_0)$ tends in probability to $\inf_{u\in\mathcal{S}'} f_0(u) > 0$, while $\sup_{u\in\mathcal{S}'} |\widehat{f}(u,\zeta_0) - f_0(u)| = o_P(1)$. Besides, $\widehat{q}(.,\zeta_0)$ is bounded by 1 and by Lemma 3.1, $\partial f/\partial\zeta(.,\zeta_0)$ is continuous on the compact set $\mathcal{S}$ and thus is bounded on this set. Thus, it suffices to prove that

$$\sup_{u\in\mathcal{S}'} \left| \frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0) - \frac{\partial f}{\partial\zeta}(u,\zeta_0) \right| = o_P(1) \qquad (3.17)$$

and similarly for r_0. By Lemma 3.1, $u \mapsto \partial f/\partial\zeta(u,\zeta_0)$ is the derivative of $-E(X|U = u)f_0(u)$. As a consequence, we can apply Newey & McFadden (1994)'s Lemma 8.10, using as before Assumptions 3.2, 3.6, and conditions on h_n. This yields (3.17). The same reasoning applies to r_0, yielding (3.16).

Now, let us establish that

$$\sup_{u\in\mathcal{S}'} \left\| \frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_{u,n}) - \frac{\partial\widehat{q}}{\partial\zeta}(u,\zeta_0) \right\| = o_P(1)$$

Using a similar decomposition as previously and the preceding results, it suffices to prove that

$$\sup_{u\in\mathcal{S}'}\left\|\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_{u,n})-\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0)\right\|=o_P(1) \tag{3.18}$$

and similarly for $\widehat{r}$. By Assumption 3.6, there exists $C_5>0$ such that $|K'(u)-K'(v)|\leq C_5|u-v|$. Thus,

$$\begin{aligned}\left\|\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_{u,n})-\frac{\partial\widehat{f}}{\partial\zeta}(u,\zeta_0)\right\| &\leq \frac{1}{nh_n^2}\sum_{i=1}^n\|X_i\|\left|K'\left(\frac{u-X_i'\zeta_{u,n}}{h_n}\right)-K'\left(\frac{u-X_i'\zeta_0}{h_n}\right)\right|\\ &\leq \frac{C_5C_0^2\left\|\zeta_{u,n}-\zeta_0\right\|}{h_n^3}=O_p\left(\frac{1}{\sqrt{n}h_n^3}\right).\end{aligned}$$

This proves (3.18) since $nh_n^6\to\infty$. The same reasoning applies to $\widehat{r}$. The result follows.

References

Carneiro, P., Hansen, K. & Heckman, J. (2003), 'Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice', *International Economic Review* **44**, 361–422.

D'Haultfoeuille, X. & Maurel, A. (2012), 'Another look at the identification at infinity of sample selection models', *Econometric Theory* **Forthcoming**.

Klein, R. W. & Spady, R. H. (1993), 'An efficient semiparametric estimator for binary response models', *Econometrica* **61**, 387–421.

Kuruscu, B. (2006), 'Training and lifetime income', *The American Economic Review* **96**, 832–846.

Newey, W. K. (2009), 'Two step series estimation of sample selection models', *The Econometrics Journal* **12**, S217–S229.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. Engle & D. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier.