Part

Online Appendix to "A Nonparametric Non-classical Measurement Error Approach to Estimating Intergenerational Mobility Elasticities"

Table of Contents

Appendices

A Monte Carlo Experiments

In the Appendix, we present Monte Carlo evidence to illustrate the finite-sample performance of the estimators in our paper and show our test's effectiveness. These results showcase the superior finite-sample performances of our proposed approach, even for modest size samples. Our designs accommodate measurement errors due to *both* self-reported *and* life-cycle bias, as well as nonlinearity.

We assume that the transitory shocks U_t and V_t are i.i.d. draws from the standard normal $\mathcal{N}(0,1)$. For illustrative purposes, we only simulate the data for two periods, and we fix $\alpha_2 = 1.10$ and $\delta_2 = 1.184$. Since our estimation procedure is conditional on Z, we need not worry about covariates in our Monte Carlo studies.

We consider three specifications of the function $g(X^*)$ and its derivative function (IGEs) $g'(X^*)$. (i) A linear mobility function and constant IGEs: $Y^* = g(X^*) + \varepsilon = 0.6X^* + \varepsilon, g'(X^*) = 0.6$. (ii) A quadratic mobility function and linear IGEs: $Y^* = g(X^*) + \varepsilon = X^* - 0.2X^{*2} + \varepsilon, g'(X^*) = 1 - 0.4X^*$. (iii) A cubic mobility function and quadratic IGEs: $Y^* = g(X^*) + \varepsilon = X^* - 0.2X^{*2} + 0.1X^{*3} + \varepsilon, g'(X^*) = 1 - 0.4X^* + 0.3X^{*2}$. In all the three specifications, X^* follows a normal distribution with mean zero and a standard deviation 2, and the structure error ε follows normal with mean zero and a standard deviation 0.1. The number of periods $T = 2$ and the sample size $N = 100, 300, 500$. We replicate each experiment for 1000 times.

In the first set of experiments, we maintain the normalization assumption $\alpha_1 = \delta_1 = 1$. We present the estimates of α_2 and δ_2 in Table A1. The results illustrate that our estimates perform very well even for a modest sample size of 300, which is smaller than the sample size used in most of the existing studies in this context; the superior finite-sample performances of our method highlight its potential use for empirical studies in this field. We present the nonparametrically estimated density of X^* in Figure A.1 for $N = 500$ under the three specifications. The results show that the estimates track the true density closely. The estimates of the derivative function $g'(\cdot)$ are presented in Figure A.2, where the linear, quadratic, and the cubic specifications are shown on the left, the middle, and the right, respectively. The figures showcase how the estimated functions closely capture the true derivative functions for both the linear and quadratic cases, and how the $[10\%, 90\%]$ point-wise confidence intervals improve significantly as the sample size increases.

Based on the estimated derivative functions, we formally test the functional form of the mobility function. The results are presented in Table A2. In all the tests, the null hypothesis is that $g'(\cdot)$ is a constant. When the true derivative is constant or linear, we use linear specification of the derivative as our alternative; when true model is quadratic, we use quadratic as our alternative. Thus the first row provides test size, and the next two rows are power.

Table A1: Estimate of α and δ

	Parameters			
	$\alpha_2 = 1.105$		$\delta_2 = 1.184$	
	Mean	Std. Dev.	Mean	Std. Dev.
Mobility function: $g(X^*) = 0.6X^*$				
$N = 100$	1.109	0.104	1.193	0.159
$N = 300$	1.108	0.057	1.187	0.087
$N = 500$	1.105	0.043	1.184	0.064
Mobility function: $g(X^*) = -0.2X^{*2} + X^*$				
$N = 100$	1.111	0.099	1.187	0.096
$N = 300$	1.109	0.055	1.185	0.052
$N = 500$	1.105	0.041	1.184	0.038
Mobility function: $g(X^*) = X^{*3} - 0.2X^{*2} + X^*$				
$N = 100$	1.109	0.092	1.184	0.008
$N = 300$	1.108	0.053	1.184	0.004
$N = 500$	1.107	0.042	1.184	0.003

The coefficients in period $t = 1$ are $\alpha_1 = \delta_1 = 1$ by construction. Standard errors are computed by using sample standard deviation of 1000 replications.

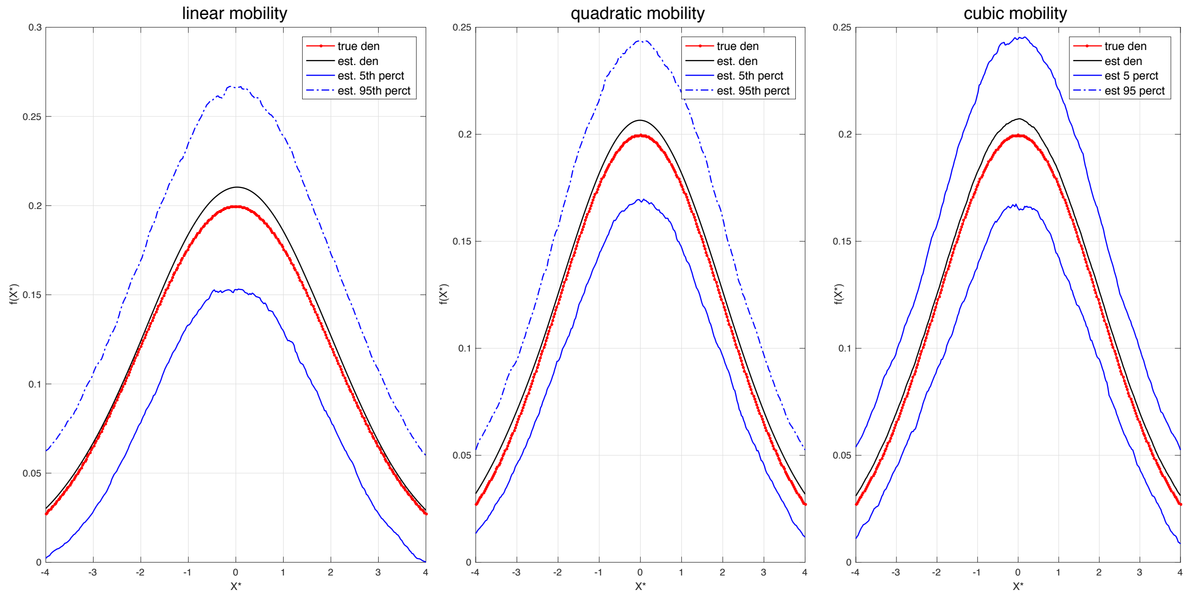
Table A2: Results of testing linearity

	$\alpha = \delta = 1$			$\alpha = \delta = 0.95$			$\alpha = \delta = 0.90$		
DGP	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$	$N = 100$	$N = 300$	$N = 500$
constant (size)	0.0%	2%	4%	0.0%	1%	4%	2%	3%	5%
linear (power)	21%	63%	65%	26%	65%	69%	29%	58%	69%
quadratic (power)	10%	32%	65%	13%	40%	63%	18%	44%	66%

Note: "constant", "linear", and "quadratic" indicate that in the date generating process, the derivative of the mobility function is constant, linear, and quadratic, respectively. The null hypothesis is $g'(\cdot)$ that is a constant. When the true derivative is constant or linear, we use linear specification of the derivative as our alternative; when true model is quadratic, we use quadratic as our alternative. Thus the first row is test size, and the next two rows are test power.

The test results in Table A2 are satisfying. We first consider the cases where the normalization is correct, i.e., $\alpha = \delta = 1$. When the true derivative function is a constant, the size of the test for $N = 500$ is 4%, very close to the nominal size 5%. When the true derivative function is linear or quadratic, the power of the test increases significantly in sample size and it's about 70% for the moderate sample size $N = 500$. In the second set of experiments, we investigate the impact of mis-normalization on the test of linearity. Specifically, we assume that the normalization assumption $\alpha_1 = \delta_1 = 1$ does not hold in the data generating process, while in estimation we still impose this assumption. The testing results for $\alpha = \delta = 0.95$ and $\alpha = \delta = 0.90$ are similar to those for $\alpha = \delta = 1$. These results, again, showcase that the normalization does not affect our testing results, as discussed in Section 3, although the incorrect normalization affects the estimates and their confidence bands.

Figure A.1: Estimated density of X^*, $N = 500$



B Further Discussions and Extensions

B.1 Continuous Variables

When the vector of covariates Z is continuous and discretizing it is not completely satisfactory in estimation, one may need to impose more structure on the mobility function $g(\cdot)$. Here, we consider a widely used partially linear model

$$Y^* = g(X^*, Z) + \epsilon = Z'\beta + h(X^*) + \epsilon, \tag{1}$$

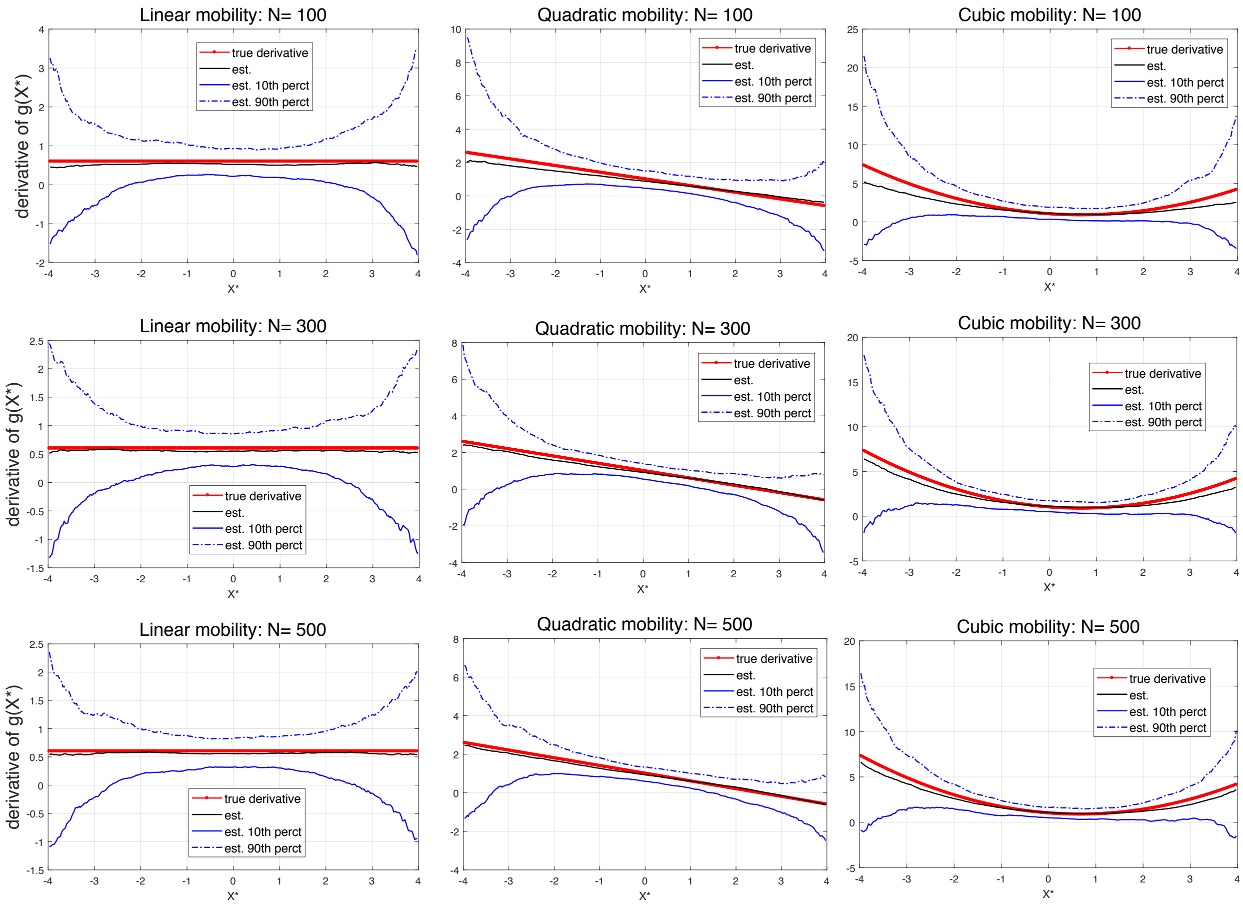
with the associations of the annual income and permanent income being

$$\begin{aligned} Y_1 &= Y^* + U_1, \\ X_1 &= X^* + V_1, \\ X_2 &= \alpha_2 X^* + V_2. \end{aligned} \tag{2}$$

Identification: For such a specification, we can identify the nonparametric function $h(\cdot)$ and the parameter β from the joint distribution of (Y_1, X_1, X_2, Z) in several steps.

First, the density of the permanent income X^*, $f_{X^*}(\cdot)$, can be identified from the joint distribution of X_1 and X_2 by applying Lemma 2. Next, the expectation of child income conditional on parental permanent income, $E(Y_1|X^*)$, can be identified using the identified density $f_{X^*}(x^*)$ together with Y_1, and X_1 based on Proposition 1. Following the similar argument, the mean of child characteristics Z

Figure A.2: Estimate of derivative functions



conditional on parental permanent income, $E(Z|X^*)$, can also be identified. It is easy to derive that

$$Y_1 - E(Y_1|X^*) = (Z' - E(Z'|X^*))\,\beta + (\epsilon + U_1)\,.$$

If we assume that the matrix $(Z' - E(Z'|X^*))'\,(Z' - E(Z'|X^*))$ is full rank, then β is identified. Once β is identified, we can identify the non-parametric function $h(X^*)$ via the following closed-form expression

$$h(X^*) = E(Y_1|X^*) - E(Z'|X^*)\beta.$$

If Z and X^* are independent, the identification above fails because $(Z' - E(Z'|X^*))'\,(Z' - E(Z'|X^*))$ is zero. If this is the case, the model can be identified as follows. $E[Y_1|z_k] = h(X^*) + z_k'\beta, k = 1, 2,$ where z_1 and z_2 are two different realizations of Z. Then $E[Y|z_1] - E[Y|z_2] = (z_1' - z_2')\beta$. Under the assumption that the matrix $(z_1' - z_2')(z_1 - z_2)$ is full rank, β is identified. Given β is identified, we can restructure the original mobility model as $Y_1 - Z'\beta = h(X^*) + \epsilon + U_1$ and identify the mobility function $h(X^*)$ using Lemma 3 in our paper.

Estimation: Note that since the identification procedure above is constructive, one can follow the identification procedure to estimate the model directly, as we do in our paper. An alternative approach to estimate $(\beta, h(\cdot))$ is discussed in Section 4.2 of Härdle et al. (2012).

B.2 Sieve MLE as an Alternative Method

We have also estimated the mobility function in the empirical application by using a sieve MLE and present the result in Figure B.3 below, along with our nonparametric estimates. Both methods indicate a U-shape in the mobility elasticities with respect to parental income. The point-wise confidence intervals for sieve MLE are larger than our estimates, however.

Below, in addition to estimation details, we discuss why, despite its nice *theoretical* properties, the sieve MLE method may not necessarily be the best alternative *in practice* for this **particular** context. We also discuss the more restrictive assumptions required for the Sieve MLE to take into account multiple measurements, and then discuss an alternative approach within our framework.

1. Sieve MLE Estimation Details: To be consistent with the data requirements of our nonparametric method, we use the joint distribution of Y, X_1, X_2 to conduct the sieve MLE. To be specific,

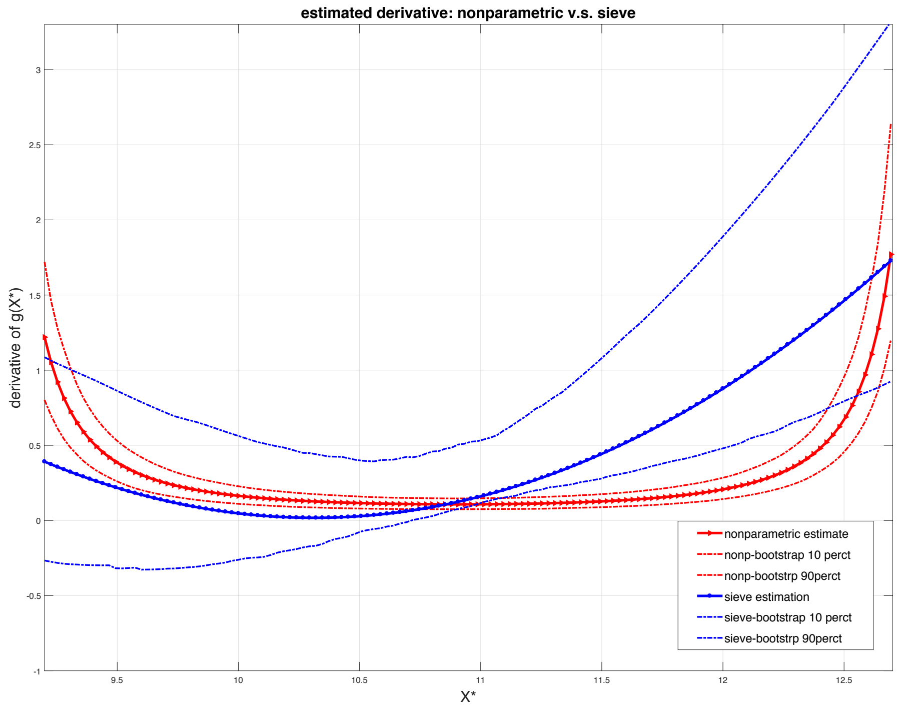
$$\begin{aligned} Y &= g(X^*) + U + \varepsilon \\ &= \beta_0 + \beta_1 x^* + \beta_2 x^{*2} + \beta_3 x^{*3} + \varepsilon' \\ X_1 &= X^* + V_1 \\ X_2 &= \alpha X^* + V_2 \end{aligned} \tag{3}$$

To proceed, we assume that Y, X_1 and X_2 are independent conditional on $X^* = x^*$. Under this assumption, the joint distribution of Y, X_1 and X_2 is

$$\begin{aligned} f_{Y,X_1,X_2}(y, x_1, x_2) &= \int f_{Y,X_1,X_2|X^*=x^*}(y, x_1, x_2|x^*) f_{X^*}(x^*) dx^* \\ &= \int f_{Y|X^*=x^*}(y|x^*) f_{X_1|X^*=x^*}(x_1|x^*) f_{X_2|X^*=x^*}(x_2|x^*) f_{X^*}(x^*) dx^* \\ &= \int f_{\varepsilon'}(y - g(x^*)) f_{V_1}(x_1 - x^*) f_{V_2}(x_2 - \alpha x^*) f_{X^*}(x^*) dx^*. \end{aligned} \tag{4}$$

We first estimate the coefficient of life-cycle bias α as we do in the paper. Next, we use sieve MLE to estimate the parameters $\beta_j, j = 0, 1, 2, 3$ together with sieve parameters for the four unknown density functions $f_{\varepsilon'}, f_{V_1}, f_{V_2}$ and f_{X^*}. We use Hermite polynomials with order four for sieve approximations.

Figure B.3: Comparison between sieve MLE and nonparametric estimation

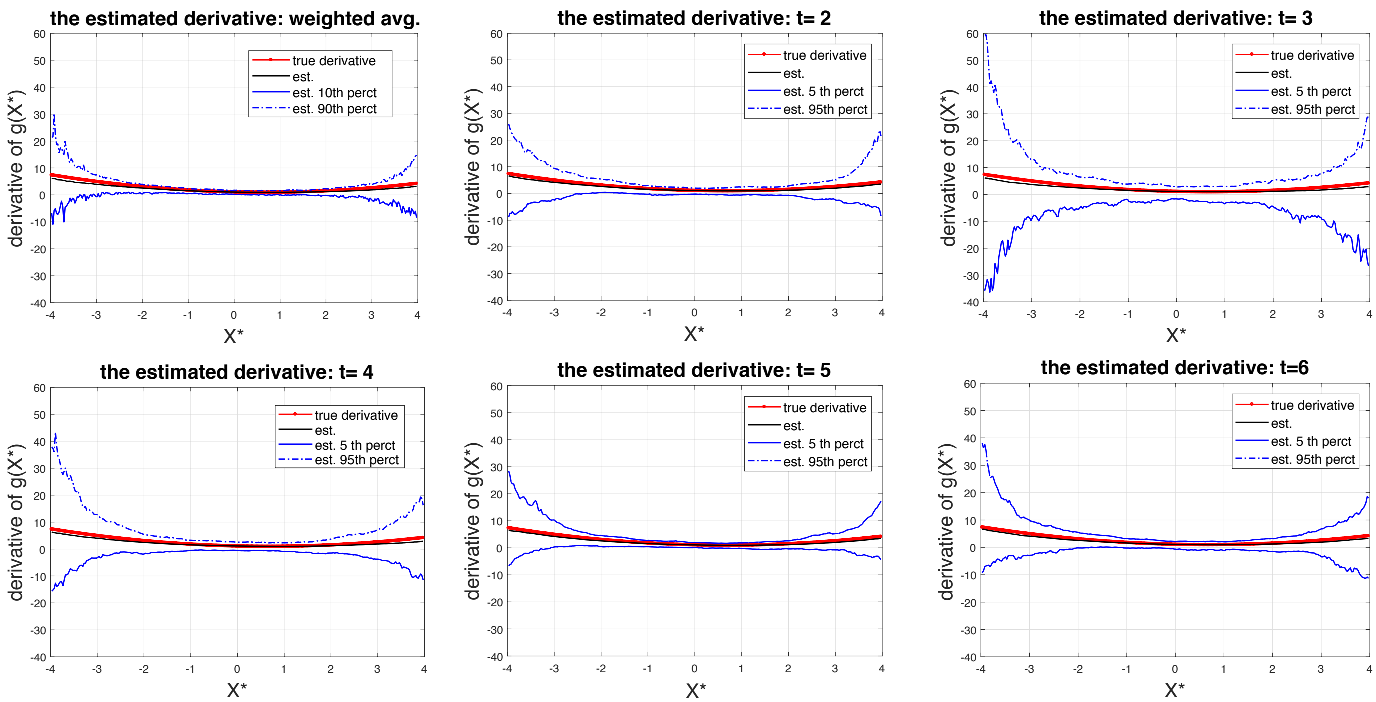


2. Further Discussions of Sieve MLE's practical issues: While sieve MLE is an excellent candidate for an alternative due to its nice theoretical properties, it has several practical disadvantages in this particular context of estimation of the mobility function. First, sieve MLE requires independence of Y, X_1, X_2 conditional on X^*, which is much more restrictive than our identification assumption. Second, our nonparametric estimation is a global one while sieve MLE is local. Possible local maximum may affect its performance in empirical applications. Finally, there is no theoretical guidance of choosing the order of the basis (Hermite polynomials in our case): the higher order it is, the smaller approximation error (and hence better performance of sieve MLE). Therefore, one faces the tradeoff between the approximation error and computational burden.

B.3 Multiple Measurement

In the last set of experiments, we investigate the possible improvement of our estimation by incorporating data of multiple periods by using model averaging. For this purpose, we first simulate data for six periods with $\alpha_1 = \delta_1 = 1, \alpha_2 = 1.10, \delta_2 = 1.184, \alpha_3 = \delta_3 = 0.8, \alpha_4 = \delta_4 = 0.9, \alpha_5 = \delta_5 = 1.2, \alpha_6 = \delta_6 = 1.05$. Then we estimate the five derivative functions using the joint distribution (Y_t, X_t, X_1) for $t = 2, \cdots, 6$. Finally, we take the weighted average of the five derivative functions through the inverse-variance weighting method. The results are presented in figure B.4, where the first plot is the weighted average and the next five ones are estimates for $t = 2, \cdots, 6$, respectively. As expected, the weighted average reduces standard errors, especially when the true income is extreme. Nevertheless, the improvement is relatively marginal, and the shape for each of the five derivative functions maintains.

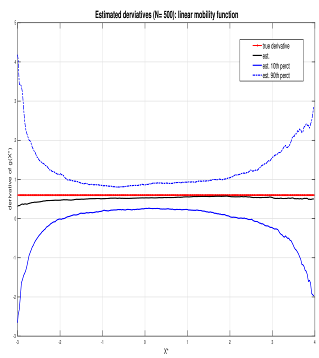
Figure B.4: Estimate of derivative functions: Weighted average



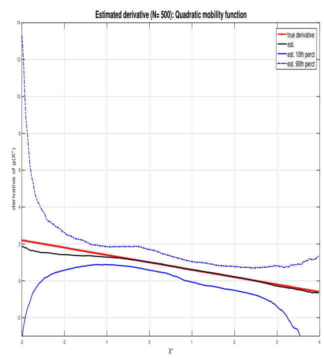
C Monte Carlo: Excluding Extremely Low Values

We also assess the robustness of our approach to exclusion of the zeros in more general settings via Monte Carlo experiments. Given our model set-up with the log income, we cannot directly simulate log zeros. Note that the log of zero is negative infinity, we therefore consider dropping the extreme values defined as bottom 5 and 10 percent of the distribution of observed (simulated) parental incomes in our simulation experiments (Figure C.5). Not surprisingly, we find that exclusion of extreme values disproportionately impacts the estimates in the lower tail, but the patterns of nonlinearity are rather robust. This may highlight the usefulness of our nonparametric approach in practice when using the administrative data.

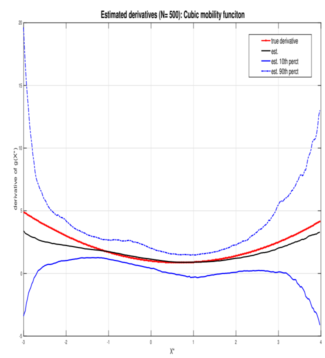
Figure C.5: Monte Carlo simulation results dropping the bottom tail of the distribution



(a) LInear

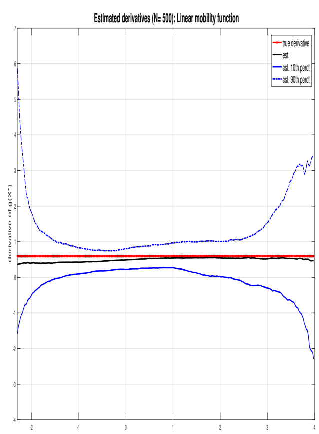


(b) Quadratic

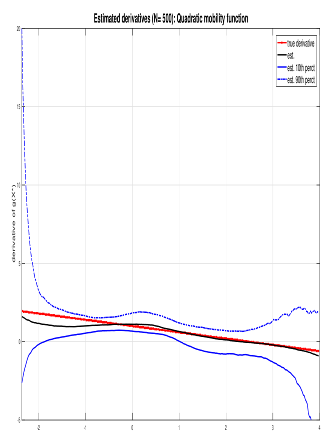


(c) Cubic

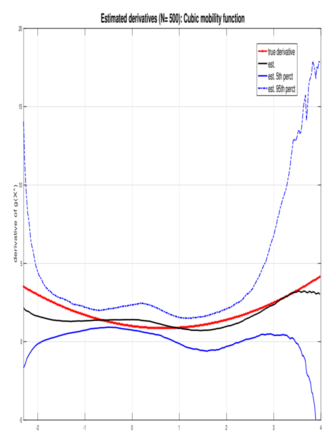
[Five Percent]



(d) Linear



(e) Quadratic



(f) Cubic

[Ten Percent]

References

Härdle, Wolfgang, Hua Liang, and Jiti Gao, *Partially linear models*, Springer Science & Business Media, 2012.