Binary Treatment Effects with Misclassification & Endogeneity

Two-step Parametric Estimation of Binary Treatment Effects with Misclassification and Endogeneity

Georgios Marios Chrysanthou

Email: gchrysanthou001@dundee.ac.uk

University of Dundee
Department of Economics

December 2024

Background

- **Binary misclassified endogenous variable**.
- Structural form: binary endogenous & misclassified explanatory variable.
- Reduced form: binary choice with misclassified dependent variable.
- Do **not** need extraneous misclassification information
- Do **not** need alternative measurements
- Approximate consistency via MMLE of misclassification probabilities & misclassification-adjusted control function
- **Contribution**:

  Bias reduction property of 2SMLS vs OLS, naive IV.

Presentation Outline

- Literature
- The Model
- Modified LS (MLS) Consistency
- Two Stage MLS Estimation (2SMLS)
- Simulations
- Empirical Application

Survey Measurement Error; Celhay, Meyer, Mittag (2024)

- Current Population (CPS), American Community (ACS), Income and Program Participation (SIPP) linked to NYS admin data on Supplemental Nutrition Assistance (SNAP), Temporary Assistance for Needy Families (TANF), General Assistance.
- Binary error: **FN** (true recipients don't report), **FP** (reported receipt but not in admin data).
- FN (18%,59%), FP (0.3%,1.3%)
- Recall error (overreporting).
- Importance (salience) $\rightarrow$ answer quality: (duration, amount).
- Stigma reduces social benefit program receipt reporting.
- Cooperativeness $\rightarrow$ response accuracy (frequently nonrespondents more likely to misreport).

Literature

- **Category 1.** Exogenous binary treatment and misclassification
  (*e.g.* Aigner, 1973; Lewbel, 2007).
- **Category 2.** Endogenous binary treatment & exogenous (non-differential) misclassification
  (*e.g.* Battistin *et al.*, 2014; Calvi *et al.*, 2022; Tommasi and Zhang, 2024b).
- **Category 3.** Endogenous binary treatment & endogenous (differential) misclassification
  (*e.g.* Ura, 2018; Nguimkeu *et al.*, 2019; Tommasi and Zhang, 2024a).
- Current study in **Category 3**:
  extends MLS (Aigner, 1973) adding endogeneity, and estimates misclassification via MMLE (Hausman *et al.*, 1998).

Misclassified Binary Endogenous Explanatory Variable

$$
\begin{aligned}
y_{j,i} &= \mathrm{z}_i \boldsymbol{\beta} + \delta x_i^T + \varepsilon_{j,i} && (1) \\
j &= \{0,1\} \text{ } if \text{ } x_i^T = \{0,1\},\ y_i = y_{j,i} \quad if \quad x_i^T = j, (i = 1, ..., N)
\end{aligned}
$$

$\boldsymbol{\beta}$ $[(k-1)\text{x}1]$, z_i $[1\text{x}(k-1)]$, $\varepsilon_{j,i} \sim iid \ N(0, \sigma_\varepsilon^2)$, $\mathbb{E}(\mathrm{z}_i' \varepsilon_{j,i}) = \mathbf{0}$.

$$
x_i^T = 1\{\mathrm{w}_i \gamma + \eta_i > 0\}, (i = 1, ..., N) \qquad (2)
$$

γ $[k\text{x}1]$; w_i $[1\text{x}k]$, $\eta_i \sim iid \ N(0,1)$.

- Structural identification $\mathrm{w}_i \neq \mathrm{z}_i$
- *i.e.* $\mathrm{w}_i = \mathrm{z}_i + \mathrm{m}_i$, m_i non-empty, not collinear with z_i

Misclassification

$$y_i = z_i\boldsymbol{\beta} + \delta x_i^T + e_i, e_i = x_i^T\varepsilon_{1,i} + (1 - x_i^T)\varepsilon_{0,i} \tag{3}$$

$$\begin{pmatrix} e_i \\ \eta_i \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{matrix} \sigma_e^2 & \sigma_{e\eta} \\ \sigma_{e\eta} & 1 \end{matrix}\right). \tag{4}$$

- (x_i^T, x_i) true/observed, $x_i = x_i^T + \tau_i$, τ_i measurement error

$$\begin{aligned} y_i &= z_i\boldsymbol{\beta} + \delta x_i + u_i, u_i = (e_i - \delta\tau_i) \\ e_i &= (x_i - \tau_i)\,\varepsilon_{1,i} + (1 - x_i + \tau_i)\varepsilon_{0,i}. \end{aligned} \tag{5}$$

- Endogenous (differential) misclassification

$$x_j \not\perp \left(y_j, x_j^T\right), j = 0, 1.$$

Misclassification Probabilities

- False negative/positive (λ_1,λ_2) depend on x_i, but not (w_i, η_i)

$$\lambda_1 = \Pr(x_i = 0 | x_i^T = 1), \lambda_2 = \Pr(x_i = 1 | x_i^T = 0) \qquad (6)$$

$$Cov(x_i, \tau_i) = (\lambda_1 + \lambda_2)\, \tilde{\pi}(1 - \tilde{\pi})$$

$$\mathbb{E}(x_i) = \tilde{\pi},\ \ Var\,(x_i) = \tilde{\pi}(1 - \tilde{\pi}), \tilde{\pi} = (N)^{-1} \sum_{i=1}^{N} x_i$$

$$\mathbb{E}(\tau_i) = \lambda_1 \tilde{\pi} - \lambda_2(1 - \tilde{\pi}), \mathbb{E}(x_i \tau_i) = \lambda_1 \tilde{\pi}$$

- $Corr(x_i^T,\ \tau_i) < 0$ ($x_i^T = 1$, τ_i=-1/0; $x_i^T = 0$, τ_i= 0/1)
- $Cov(x_i, \tau_i) \neq 0$ bias unless covariates orthogonal to x_i.

Modified LS (MLS)

- Following Aigner (1973) the partitioned linear model for $(\widehat{\beta}, \widehat{\delta})_{MLS}$ is:

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\delta} \end{bmatrix} = \begin{bmatrix} M_{ZZ} & M_{Zx} \\ M'_{Zx} & (m_{xx} - \xi) \end{bmatrix}^{-1} \begin{bmatrix} M_{Zy} \\ m_{xy} \end{bmatrix} \quad (7)$$

$$M_{ZZ} = (N)^{-1} Z'Z, M_{Zx} = (N)^{-1} Z'x, M_{Zy} = (N)^{-1} Z'y$$

$$m_{xx} = (N)^{-1} x'x, m_{xy} = (N)^{-1} x'y, \xi = Cov(x_i, \tau_i).$$

where, $Z_{Nx(K-1)}$.

Consistency MLS Estimation

- Partitioned linear model LS normal equations

$$\begin{bmatrix} M_{ZZ} & M_{Zx} \\ M'_{Zx} & (m_{xx} - \xi) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \delta \end{bmatrix} = \begin{bmatrix} M_{Zy} \\ m_{xy} \end{bmatrix} \tag{8}$$

$$M_{Zy} = M_{ZZ}\boldsymbol{\beta} + M_{Zx}\delta, m_{xy} = M'_{Zx}\boldsymbol{\beta} + (m_{xx} - \xi)\,\delta$$

$$\delta = \left[m_{xx}^{-1} m_{xy} - m_{xx}^{-1} M'_{Zx}\boldsymbol{\beta}\right] \left[1 - \xi m_{xx}^{-1}\right]^{-1}$$

$$\widehat{\boldsymbol{\beta}} = (Z'AZ)^{-1} Z'Ay, A = \left[I - x m_{xx}^{-1} (N)^{-1} x' \left(1 - \xi m_{xx}^{-1}\right)^{-1}\right] \tag{9}$$

$$\widehat{\delta} = (N)^{-1} m_{xx}^{-1} x' \left[I - Z(Z'AZ)^{-1} Z'A\right] y \left[1 - \xi m_{xx}^{-1}\right]^{-1}. \tag{10}$$

Consistency MLS Estimation

Theorem (1)

MLS gives $\underset{N \longrightarrow \infty}{plim} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$,

$$\underset{N \longrightarrow \infty}{plim} \widehat{\delta} = \delta + \underset{N \longrightarrow \infty}{plim} \left[m_{xx}^{-1} \left(\frac{1}{N} x' e \right) \left[1 - \xi m_{xx}^{-1} \right]^{-1} \right] \tag{11}$$

corresponding to

$$\underset{N \longrightarrow \infty}{plim} \widehat{\delta} = \delta + \underset{N \longrightarrow \infty}{plim} \left[\left(x' x \right)^{-1} \left(x' e \right) \psi \right], \psi = \left[1 - \xi m_{xx}^{-1} \right]^{-1}. \tag{12}$$

Estimation

Using Eq.(3) take expectations conditioning on (w_i, x_i):

$$\mathbb{E}[y_i|\mathrm{w}_i, x_i] = \mathrm{z}_i\boldsymbol{\beta} + \delta x_i + \mathbb{E}[e_i|\mathrm{w}_i, x_i] - \delta\mathbb{E}[\tau_i|\mathrm{w}_i, x_i], \mathbb{E}[\tau_i|\mathrm{w}_i, x_i] = \xi$$

$$\mathbb{E}[y_i|\mathrm{w}_i, x_i] = \mathrm{z}_i\boldsymbol{\beta} + \delta\left(x_i - \xi\right) + \mathbb{E}[e_i|\mathrm{w}_i, x_i]. \tag{13}$$

- (e_i, η_i) *iid* Normal, **Theorem (1)**:

$$\underset{N\longrightarrow\infty}{plim}\ \widehat{\delta} = \delta + \underset{N\longrightarrow\infty}{plim}\left[\left(x'x\right)^{-1}\left(x'e\right)\psi\right], \psi = \left[1 - \xi m_{xx}^{-1}\right]^{-1}$$

$$\mathbb{E}[e_i|\mathrm{w}_i, x_i] = \frac{\sigma_{e\eta}}{\sigma_\eta^2}\psi\left[\mu_i\left(\mathrm{w}_i\boldsymbol{\gamma}\right)\right], \psi = \left(1 - \lambda_1 - \lambda_2\right)^{-1} \tag{14}$$

$$\begin{aligned}\psi &= \left[1 - \xi m_{xx}^{-1}\right]^{-1}, m_{xx} = (N)^{-1}\left(x'x\right)\\ \xi &= \left(\lambda_1 + \lambda_2\right)\tilde{\pi}(1 - \tilde{\pi}), \tilde{\pi} = (N)^{-1}\sum_{i=1}^{N} x_i\end{aligned}$$

First Stage, Modified MLE (MMLE) Estimation

Hausman et al.(1998):

$$\ln(L^m) = \sum_{i=1}^{N} \{x_i \ln[\lambda_2 + (1-\lambda_1-\lambda_2)\Phi(\mathrm{w}_i\gamma)] + (1-x_i)\ln[1-\lambda_2-(1-\lambda_1-\lambda_2)\Phi(\mathrm{w}_i\gamma)]\} \quad (15)$$

- $\frac{\partial \ln(L^m)}{\partial \gamma}$=0 gives

$$\mu_i(\mathrm{w}_i\gamma) = \frac{\phi(\mathrm{w}_i\gamma)(1-\lambda_1-\lambda_2)[x_i-\lambda_2-(1-\lambda_1-\lambda_2)\Phi(\mathrm{w}_i\gamma)]}{[\lambda_2+(1-\lambda_1-\lambda_2)\Phi(\mathrm{w}_i\gamma)][1-\lambda_2-(1-\lambda_1-\lambda_2)\Phi(\mathrm{w}_i\gamma)]}$$

where $\eta_i \sim N(0, \sigma_\eta^2)$, (ϕ, Φ) Normal *pdf* and *cdf*.

Second Stage, MLS Estimation

- If $\lambda_1 + \lambda_2 < 1$, $(\gamma, \lambda_1, \lambda_2)$ estimated via MMLE of L^m.

Theorem (2)

Consistent LS estimation of $(\boldsymbol{\beta}, \delta, \varkappa)$ *via*

$$y_i = \mathrm{z}_i \boldsymbol{\beta} + \delta \left(x_i - \xi \right) + \left[\frac{\mu_i \left(\mathrm{w}_i \gamma \right)}{\left(1 - \lambda_1 - \lambda_2 \right)} \right] \varkappa + \omega_i \qquad (16)$$

Simulation Results Summary

- **Bias reduction** property of 2SMSLS vs OLS, naive IV.
- 2SMLS significantly outperforms OLS and can be approximately asymptotically consistent **if δ sign is opposite to $\rho = \text{corr}\left(e_i, \eta_i\right)$ sign**.
- **If δ and $\rho = \text{corr}\left(e_i, \eta_i\right)$ have the same sign, endogeneity bias and misclassification attrition bias may approximately cancel out if**:

$$0.5 \leq \left| \frac{corr(e_i, \eta_i)}{\lambda_1 + \lambda_2} \right| \leq 1$$

such that OLS bias can be lower/comparable to 2SMLS.

Endogeneity & misclassification attrition bias cancellation

- In [0.5-1] "bias cancellation region", 2SMLS bias can be reduced at approximately the same/lower level of bias compared to OLS by using inverted misclassification rates such that the IMR becomes

$$\mu_i\left(\mathrm{w}_i\gamma\right)^I = \frac{\phi\left(.\right)\left(1-\lambda_1-\lambda_2\right)^{-1}\left[x_i-\left(\lambda_2\right)^{-1}-\left(1-\lambda_1-\lambda_2\right)^{-1}\Phi\left(.\right)\right]}{\left[\lambda_2^{-1}+\left(1-\lambda_1-\lambda_2\right)^{-1}\Phi\left(.\right)\right]\left[1-\left(\lambda_2\right)^{-1}-\left(1-\lambda_1-\lambda_2\right)^{-1}\Phi\left(.\right)\right]}$$

- The above IMR adjustment (using inverted misclassification rates) is used in the second occurrence of the bold indicated simulation cases.

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.1$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho/M\rvert$	$\lvert ratio\rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
10/2	5.00	0.0771	0.1344	0.0858	0.0708	-0.3407	0.0881
10/5	2.00	0.0920	0.1035	0.1138	0.0829	-0.3592	0.1085
10/10	1.00	0.1067	0.0523	0.1543	0.1035	-0.3885	0.1535
10/10	1.00	0.0478	0.0523	0.1543			
10/20	0.50	0.1371	-0.0458	0.2485	0.1240	-0.4454	0.2512
10/20	0.50	-0.0492	-0.0458	0.2485			
10/30	0.33	0.1390	-0.1849	0.4320	0.1258	-0.5257	0.4253
10/40	0.25	0.1129	-0.3152	0.6650	0.0992	-0.6013	0.6685
10/50	0.20	0.0504	-0.4379	0.9993	0.0315	-0.6738	1.0034
10/60	0.17	-0.0543	-0.6026	1.7816	-0.0743	-0.7693	1.7857

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.2$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho / M\rvert$	$\lvert ratio \rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
20/2	10.00	0.0809	0.3676	0.0858	0.0661	-0.5738	0.0881
20/5	4.00	0.0966	0.3302	0.1136	0.0782	-0.5859	0.1087
20/10	2.00	0.1113	0.2671	0.1543	0.0989	-0.6035	0.1535
20/20	1.00	0.1414	0.1502	0.2485	0.1191	-0.6414	0.2512
20/30	0.67	0.1487	-0.0168	0.4317	0.1239	-0.6936	0.4255
20/30	0.67	-0.0107	-0.0168	0.4317			
20/40	0.50	0.1094	-0.1741	0.6680	0.0850	-0.7423	0.6680
20/50	0.40	0.0586	-0.3225	0.9992	0.0233	-0.7893	1.0034
20/60	0.33	-0.0442	-0.5209	1.7790	-0.0796	-0.8509	1.7853

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.3$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho / M\rvert$	$\lvert ratio \rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
30/2	15.00	0.0852	0.6207	0.0862	0.0601	-0.8270	0.0882
30/5	6.00	0.0999	0.5761	0.1121	0.0748	-0.8317	0.1103
30/10	3.00	0.1164	0.5023	0.1543	0.0938	-0.8385	0.1535
30/20	1.50	0.1476	0.3630	0.2485	0.1129	-0.8542	0.2512
30/30	1.00	0.1529	0.1658	0.4314	0.1119	-0.8762	0.4302
30/40	0.75	0.1256	-0.0213	0.6671	0.0814	-0.8951	0.6690
30/40	0.75	-0.0069	-0.0213	0.6671			
30/50	0.60	0.0676	-0.1971	0.9992	0.0143	-0.9146	1.0035
30/60	0.50	-0.0329	-0.4323	1.7790	-0.0908	-0.9395	1.7854

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.4$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho/M\rvert$	$\lvert ratio \rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
40/2	20.00	0.0900	0.8924	0.0858	0.0570	-1.0986	0.0882
40/5	8.00	0.1049	0.8400	0.1131	0.0698	-1.0956	0.1092
40/10	4.00	0.1218	0.7559	0.1542	0.0884	-1.0921	0.1535
40/20	2.00	0.1542	0.5912	0.2484	0.1063	-1.0824	0.2513
40/30	1.33	0.1639	0.3611	0.4311	0.1001	-1.0715	0.4262
40/40	1.00	0.1340	0.1425	0.6650	0.0730	-1.0590	0.6686
40/50	0.80	0.0772	-0.0628	0.9992	0.0047	-1.0490	1.0035
40/50	0.80	-0.0476	-0.0628	0.9992			
40/60	0.67	-0.0208	-0.3373	1.7789	-0.1029	-1.0345	1.7854

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.5$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho / M\rvert$	$\lvert ratio\rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
50/2	25.00	0.0955	1.2092	0.0883	0.0515	-1.4154	0.0855
50/5	10.00	0.1134	1.1475	0.1119	0.0641	-1.4035	0.1109
50/10	5.00	0.1282	1.0478	0.1542	0.0820	-1.3840	0.1537
50/20	2.50	0.1620	0.8574	0.2484	0.0985	-1.3486	0.2513
50/30	1.67	0.1696	0.5886	0.4269	0.0952	-1.2990	0.4310
50/40	1.25	0.1465	0.3333	0.6670	0.0656	-1.2499	0.6661
50/50	1.00	0.0884	0.0940	0.9986	-0.0065	-1.2058	1.0029
50/60	0.83	-0.0067	-0.2265	1.7724	-0.1170	-1.1454	1.7877

Relative Bias $[\frac{\hat{\delta}-\delta}{\delta}]$, $\rho = \pm 0.6$, (N=15000, R=3000)

	$\delta > 0$	$\rho > 0$			$\rho < 0$		
	$\delta < 0$	$\rho < 0$			$\rho > 0$		
$\lvert\rho / M\rvert$	$\lvert ratio \rvert$	2SMLS	OLS	IV	2SMLS	OLS	IV
60/2	30.00	0.1070	1.6167	0.0857	0.0395	-1.8230	0.0882
60/5	12.00	0.1182	1.5437	0.1124	0.0543	-1.7994	0.1099
60/10	6.00	0.1438	1.4256	0.1542	0.0656	-1.7619	0.1536
60/20	3.00	0.1771	1.1996	0.2484	0.0799	-1.6908	0.2513
60/30	2.00	0.1824	0.8811	0.4264	0.0808	-1.5917	0.4316
60/40	1.50	0.1628	0.5803	0.6669	0.0460	-1.4968	0.6692
60/50	1.20	0.1029	0.2956	0.9991	-0.0209	-1.4074	1.0036
60/60	1.00	0.0114	-0.0839	1.7788	-0.1351	-1.2879	1.7856

HH Social Benefit Income & Labour Market Inactivity

- Understanding Society, UKHLS, (wave 1, 2009), ages 25-65.
- 24 month UKHLS survey fieldwork period, individuals interviewed at approx 12 month intervals, and individual responses collected once per wave.
- Interview dates equally split between 2009, 2010 (3.35% interviews in 2011).
- Binary inactivity variable likely to be misclassified: stated current economic activity on interview date may **not** accurately reflect individual economic activity throughout the whole wave.
- Model for inactivity/unemployment: individual observations with large linear index $(\mathrm{w}_i\gamma)$ values will predict 1 with probability $\Phi\left(\mathrm{w}_i\gamma\right)$ close to 1 (and 0 for small index values) independently of observed stated individual current economic activity responses.

HH Social Benefit Income & Labour Market Inactivity

- Dependent variable: **total household social benefit income (z) month before interview**
- inverse hyperbolic sine transformation, $ln(y + \sqrt{(y^2 + 1)})$
- "Total household social benefit income" corresponds to total income aid received by UK government as part of the "Universal Credit" contains up to 39 components including pension income, incapacity benefit, income support, job seeker's allowance, child benefit, maternity allowance, housing benefit, and council tax benefit.

HH Social Benefit Income & Labour Market Inactivity

- **Binary treatment: "inactive"**
  0 *if employed/self-employed;*
  1 *if unemployed/economically inactive*
- unemployed/economically inactive*: unemployed, retired, maternity leave, family care, full-time student, long-term sick/disabled, Government training scheme, other*
- **Structural identification**: binary indicator for highest frequency of self-reported **Current Financial State**.
- "Doing Alright Financially" if response to current financial state is "Doing Alright", 0 if "Finding it very difficult", "Finding it quite difficult", "Just about getting by/don't know", "Living comfortably".

Estimation of Error Correlation $\rho = \text{corr}(e_i, \eta_i)$

- Binary inactivity/unemployment treatment indicator likely to be both **misclassified and endogenous**.
- Unobserved economic activity individual determinants of economic activity may be correlated with unobserved household benefit income determinants.
- MMLE estimation, $[\mu_{ij}(\mathrm{w}_i \hat{\gamma})]$
- Following Heckman (1979)

$$\widehat{\sigma} = \sqrt{\frac{\hat{\mathrm{e}}'\hat{\mathrm{e}}}{N}}$$

$$\widehat{\rho} = \frac{\hat{\varkappa}}{\widehat{\sigma}}$$

Table: Descriptive Statistics

	mean	standard deviation
Unemployed/Economically inactive	0.3058	(0.0363)
Doing Alright financially	0.3150	(0.0333)
Age	44.8965	(0.0017)
Female	0.5683	(0.0313)
Married/Civil Partnership	0.5734	(0.0341)
Number of children in household	0.5856	(0.0216)
LT illness/disability	0.3531	(0.0339)
University degree	0.2532	(0.0365)
House owned outright/mortgage	0.6918	(0.0370)
North West	0.1207	(0.0837)
Yorkshire and the Humber	0.0880	(0.0879)
East Midlands	0.0812	(0.0891)
West Midlands	0.0910	(0.0873)
East of England	0.0985	(0.0863)
London	0.1044	(0.0859)
South East	0.1398	(0.0821)
South West	0.0895	(0.0876)
Wales	0.0521	(0.0978)
Scotland	0.0883	(0.0878)
arsinh(household social benefit income)	4.3720	(3.2670)
Number of Observations	26896	

Table: Probability of Unemployment/Labour Inactivity, MMLE

Doing Alright Financially	-0.3165***
	(0.0343)
Age	0.0482***
	(0.0036)
Female	0.5381***
	(0.0396)
Married/Civil Partnership	-0.1475***
	(0.0296)
Number of children in household	0.2727***
	(0.0260)
LT illness/disability	0.6261***
	(0.0418)
University degree	-0.3907***
	(0.0437)
House owned outright/mortgage	-1.0952***
	(0.0797)
Constant	-2.3526***
	(0.1543)
$\widehat{\lambda_1}$	0.1964***
	(0.0290)
$\widehat{\lambda_2}$	0.0637***
	(0.0083)
Number of Observations	26896

1. Standard errors in parentheses; 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table: Total Household Social Benefit Income, UKHLS 2009

	2SMLS	OLS	IV
Unemployed/Economically Inactive	3.1210***	2.0526***	3.8808***
	(0.6454)	(0.0362)	(0.5523)
Modified Generalised Residual	-0.7222*		
	(0.3988)		
Age	0.0411***	0.0509***	0.0341***
	(0.0062)	(0.0017)	(0.0054)
Female	0.1286	0.2457***	0.0454
	(0.0789)	(0.0313)	(0.0686)
Married/Civil Partnership	-0.0275	-0.0685**	0.0016
	(0.0420)	(0.0341)	(0.0414)
Number of children in household	2.1456***	2.1972***	2.1088***
	(0.0374)	(0.0216)	(0.0350)
LT illness/disability	0.3184***	0.4772***	0.2056**
	(0.0941)	(0.0339)	(0.0892)
University degree	-0.6569***	-0.7333***	-0.6026***
	(0.0565)	(0.0365)	(0.0549)
House owned outright/mortgage	-0.4399***	-0.6934***	-0.2597*
	(0.1499)	(0.0370)	(0.1363)
Constant	1.0352***	0.8309***	0.8856***
	(0.1367)	(0.1039)	(0.1099)
$\widehat{rho} = \widehat{corr}\,(e_i, \eta_i)$	-0.2879		
$\left\lvert \frac{\widehat{\rho}}{\widehat{\lambda_1} + \widehat{\lambda_2}} \right\rvert$	1.1072		
Number of Observations	26896	26896	26896

Policy Relevance

- **Semi-elasticity** of social benefit income wrt changes in unemployment/labour market inactivity incidence.
- (2SMLS, OLS) semi-elasticities (0.9842, 0.6549).
- Demand for social benefits as per 2SMLS is 0.5029% more responsive to rises in unemployment and labour market inactivity incidence (**twice as elastic**).

ivbounds (Lin, Tommasi, Zhang; 2021, 2024)

- Tommasi and Zhang (2024a) **TZIV** estimator- **ivbounds** (Stata). "Doing Alright Financially" binary indicator to instrument "Unemployed/Economically Inactive".
- **TZIV** bounds treatment; point estimate if $(\widehat{\lambda_1}, \widehat{\lambda_2})$ known.
- **TZIV mismeasured IV estimand**: 4.116, 95% CI $[2.5313, 5.7006]$.
- **TZIV** treatment effect point estimate [**MMLE** $(\widehat{\lambda_1}, \widehat{\lambda_2})$]: $4.116(1-(\widehat{\lambda_1}+\widehat{\lambda_2})) = 4.116(0.7399) = \mathbf{3.0454}$.
- **2SMLS** point estimate **3.1210**, 95% CI $[2.7898, 3.4523]$, **2SMLS** 100-bootstrapped 95% CI $[1.8561, 4.386]$.
- **TZIV** 100-bootstrapped 95% CI $[1.873, 4.2181]$.

Concluding Remarks

- 2SMLS **bias reduction** vs OLS, naive IV (**ivregress**, Stata); approximately asymptotically consistent if δ and $\rho = \text{corr}\left(e_i, \eta_i\right)$ have opposite signs.
- If δ, $\rho = \text{corr}\left(e_i, \eta_i\right)$ sign is same, endogeneity and misclassification attrition biases can cancel out if

$$0.5 \leq \left| \frac{corr\left(e_i, \eta_i\right)}{\lambda_1 + \lambda_2} \right| \leq 1$$

  such that OLS bias is lower or comparable to 2SMLS.
- 2SMLS similar to TZIV, Tommasi and Zhang (2024a).
- TZIV (**ivbounds**, Stata) requires external misclassification probability knowledge to improve estimated treatment effect bounds.