Appendix of: "What do instrumental variable models deliver with discrete dependent variables?"

Andrew Chesher and Adam M Rosen

CeMMAP and University College London

This Appendix contains:

1. examples of identified sets delivered by an instrumental variable threshold crossing model for a binary outcome,
2. displays of conditional probability distributions of structural equation errors given endogenous explanatory variables using some data analysed in Angrist and Evans (1998).

1 Examples of identified sets delivered by the IV model

We illustrate the identified sets delivered by the IV threshold crossing model using probability distributions created by us, generated using a Gaussian triangular structure, one of the cases studied in Heckman (1978), as follows.

$$Y_1 = 1[-\alpha_\Delta - \beta_\Delta Y_2 < U^*] \qquad Y_2 = 1[-\gamma - \delta Z < V^*]$$

$$\begin{bmatrix} U^* \\ V^* \end{bmatrix} | Z \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \right) \qquad Z \in \mathcal{R}_Z \equiv \{-1, 0, 1\}$$

Here $\rho_0 = \Phi(\alpha_\Delta)$ and the Average Treatment Effect (ATE) is $\rho_1 = \Phi(\beta_\Delta + \alpha_\Delta) - \Phi(\alpha_\Delta)$ where Φ is the standard normal distribution function.

Figures 1-9 show identified sets (grey) for (ρ_0, ρ_1) at a variety of parameter settings. Throughout $\alpha_\Delta = \gamma = 0$ and $\beta_\Delta = -1$.

The value of (ρ_0, ρ_1) in the triangular Gaussian structure is marked in green. The points marked red and blue are discussed shortly. The IV model is partially identifying so other structures could have generated the probability distributions that delivered these sets. Examples include non-triangular structures and non-Gaussian structures and in those cases the value of (ρ_0, ρ_1) would be at different points in the identified set.

In this simple case the sets comprise either one, or two disconnected, parallelograms. When the endogenous variable (here binary) has K points of support identified sets are unions of up to $K!$ convex polytopes which may be disconnected, see Chesher (forthcoming) for an illustration. In Figures 1-5 the identified sets are connected, the strength of the instrument steadily increases (the coefficient δ is increasing) and as it does the identified set shrinks. There would be point identification if δ increased without limit. In Figures 6-9 the correlation r is positive and the identified set is disconnected until the instrument becomes quite strong.

The points marked red and blue in Figures 1-9 show the 2SLS and OLS estimands

$$\beta_{2sls} = \frac{cov(Z, Y_1)}{cov(Z, Y_2)} \qquad \alpha_{2sls} = E[Y_1] - \beta_{2sls} E[Y_2], \tag{1.1}$$

$$\beta_{ols} = \frac{cov(Y_1, Y_2)}{var(Y_2)} \qquad \alpha_{ols} = E[Y_1] - \beta_{ols} E[Y_2], \tag{1.2}$$

in red and blue, respectively, obtained using a linear probability model and 500 randomly chosen distributions of the instrumental variable.[1]

[1] The instrumental variable has support on $\{-1, 0, 1\}$. To generate a random distribution on this support, three independent uniform pseudo-random variates (B_1, B_2, B_3) were generated and probabilities generated as $B_i / \Sigma_j B_j$ for $i \in \{1, 2, 3\}$.

The triangular Gaussian model delivers distributions of (Y_1, Y_2) conditional on $Z \in \{-1, 0, 1\}$ and the moments appearing in (1.1) and (1.2) in general depend on the distribution of probability mass on the support of Z.

The OLS estimand is quite sensitive to the distribution of the instrumental variable. The 2SLS estimand is less variable and is often close to the triangular model value of the ATE but sometimes it falls outside the identified set. When the identified set is large it is of course far from many of the observationally equivalent values of the ATE.

2 Distribution of latent U conditional on endogenous explanatory variable Y_2

In this binary endogenous variable case, associated with each point in the identified set for (ρ_0, ρ_1) there is a unique pair of conditional distributions of U given $Y_2 = y_2 \in \{0, 1\}$ that generates the probability distributions of (Y_1, Y_2) given Z that deliver the identified set. The extent to which these differ at some value (ρ_0, ρ_1) indicates the degree of endogeneity of the explanatory variable Y_2 at that value.

Figures 10-19 show some of these distributions for the Angrist and Evans (1998) analysis (described in the main text) using the same-sex instrument. Figure 10 shows the distributions at their 2SLS estimate. There is hardly any endogeneity here. The remaining Figures show the distributions (right hand panes) associated with some extreme points in identified set (the points are highlighted in the left hand panes). Different points in the identified set are associated with widely varying degrees and directions of endogeneity.

References

Angrist, J., and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 88(3), 450–487.

Chesher, A. (forthcoming): "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *Econometric Theory*.

Heckman, J. J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

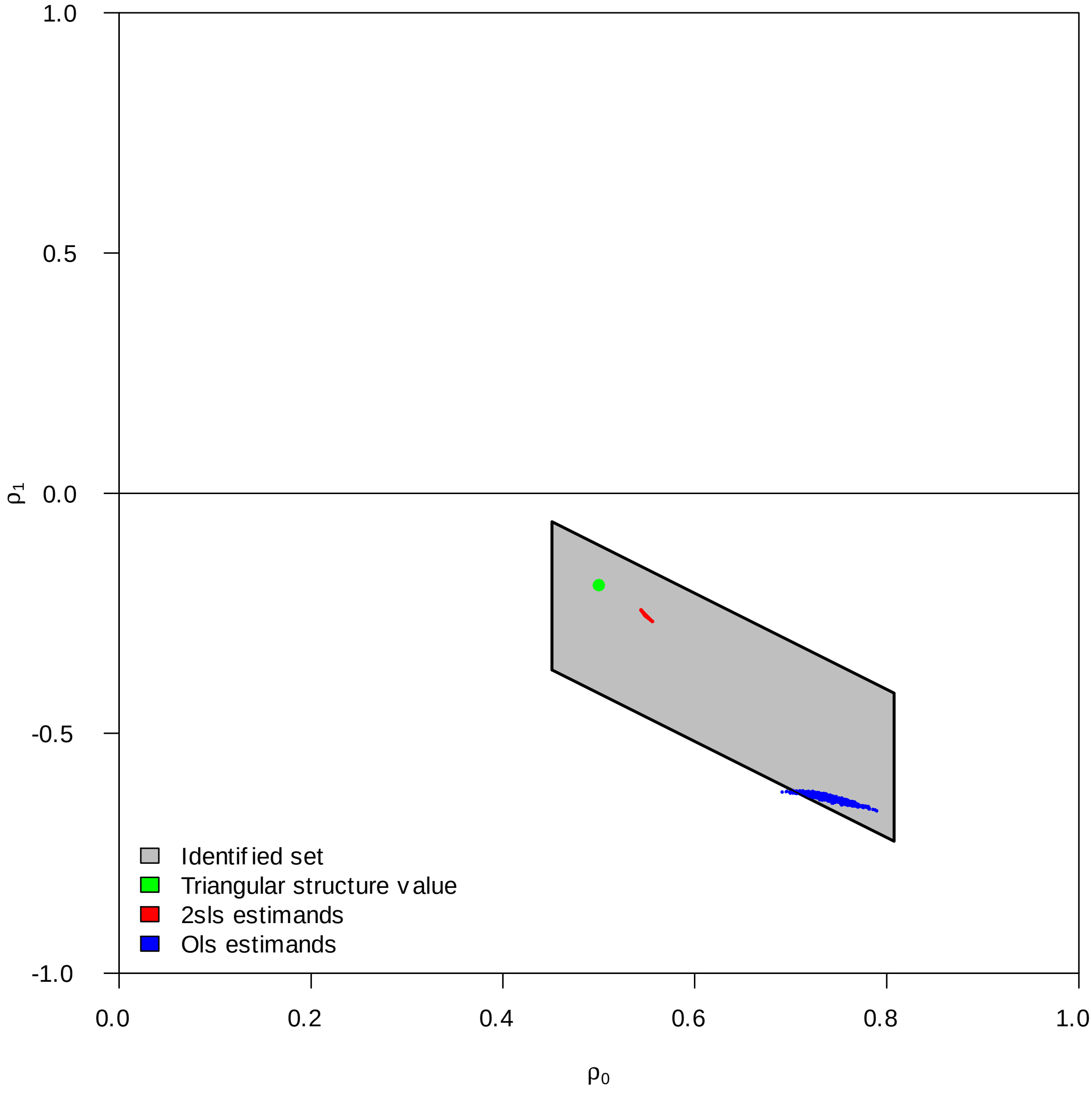

Figure 1: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = -0.7$, $\delta = 0.3$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

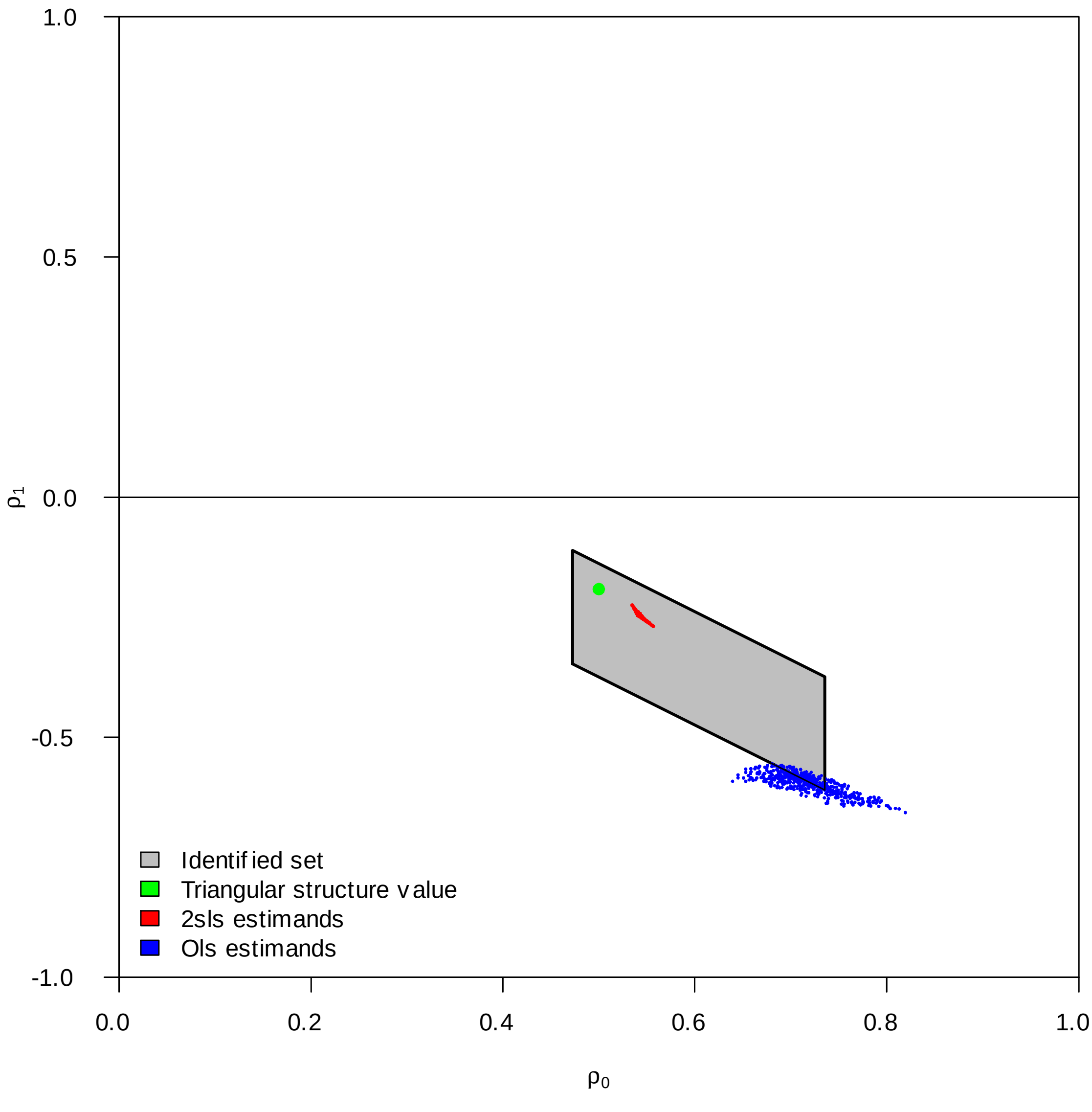

Figure 2: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = -0.7$, $\delta = 0.6$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

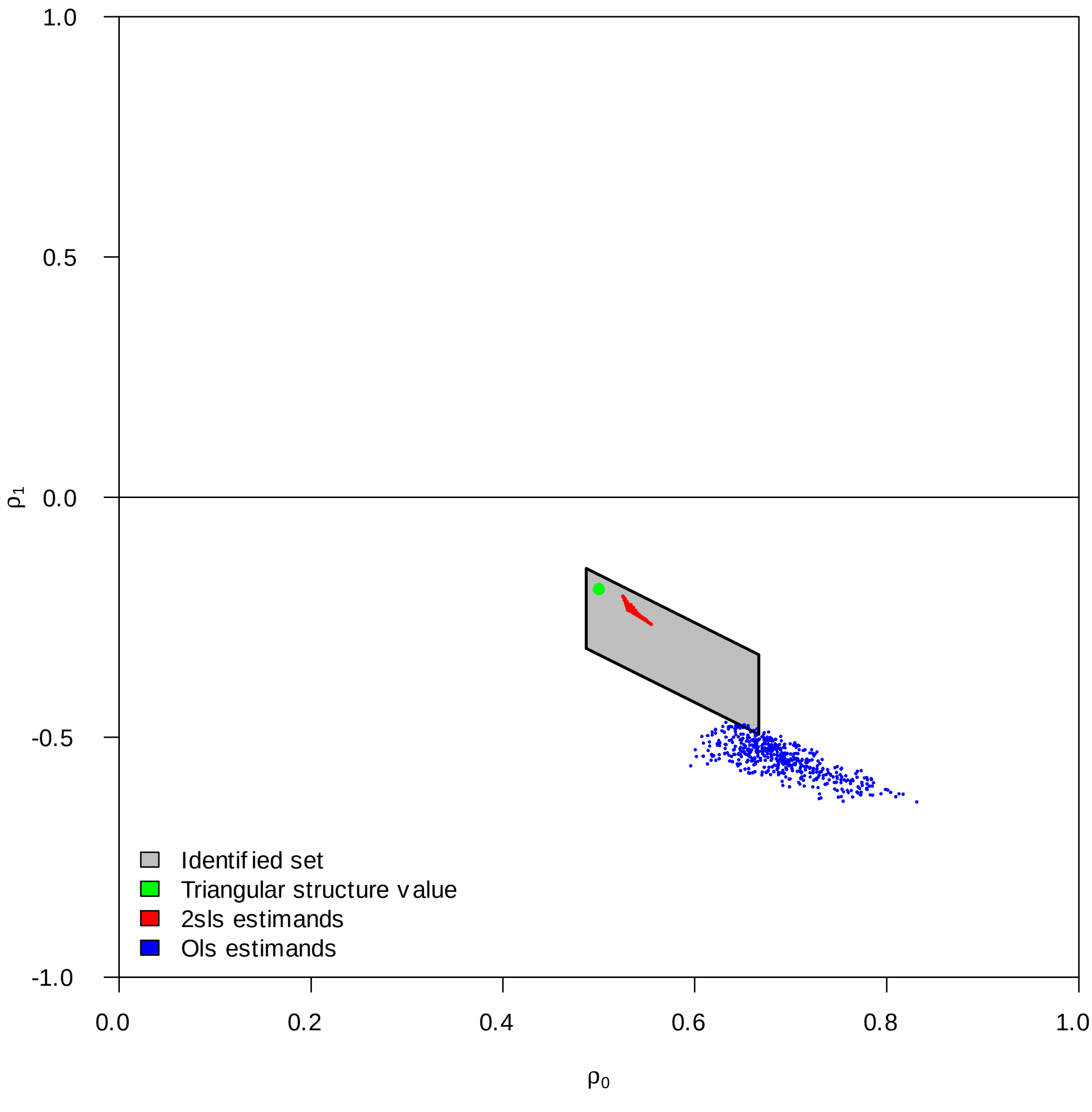

Figure 3: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = -0.7$, $\delta = 0.9$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

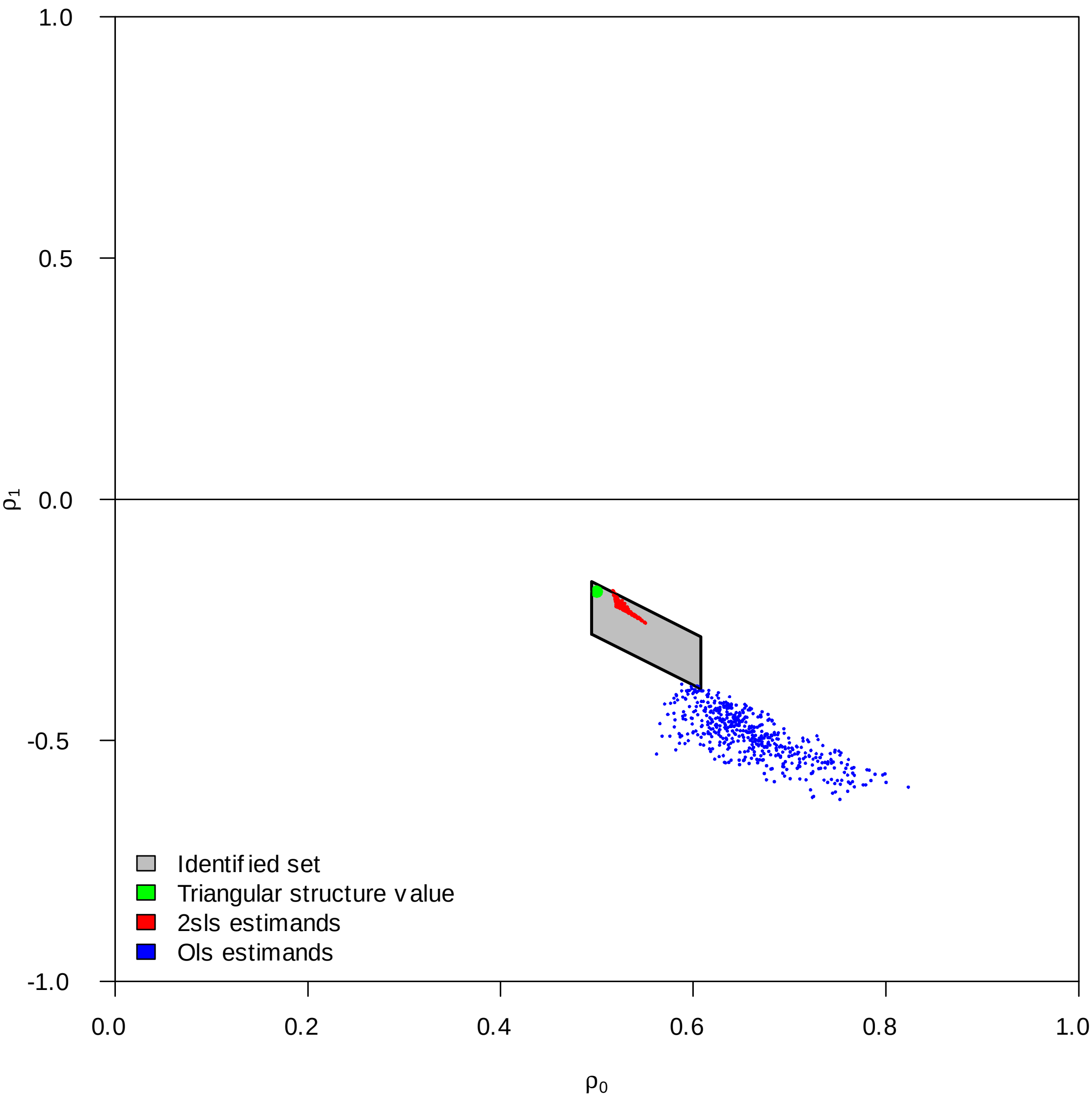

Figure 4: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = -0.7$, $\delta = 1.2$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

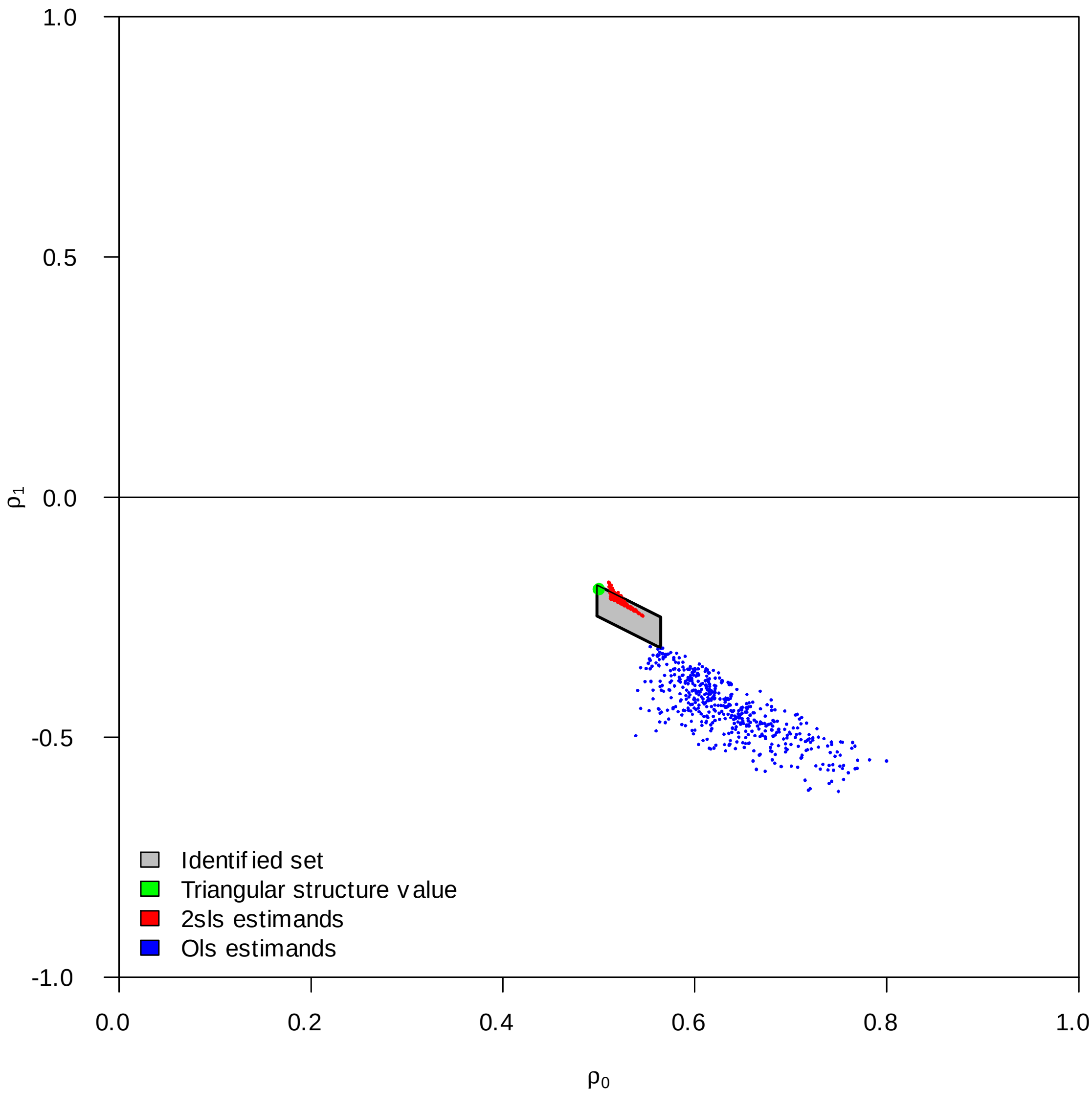

Figure 5: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = -0.7$, $\delta = 1.5$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

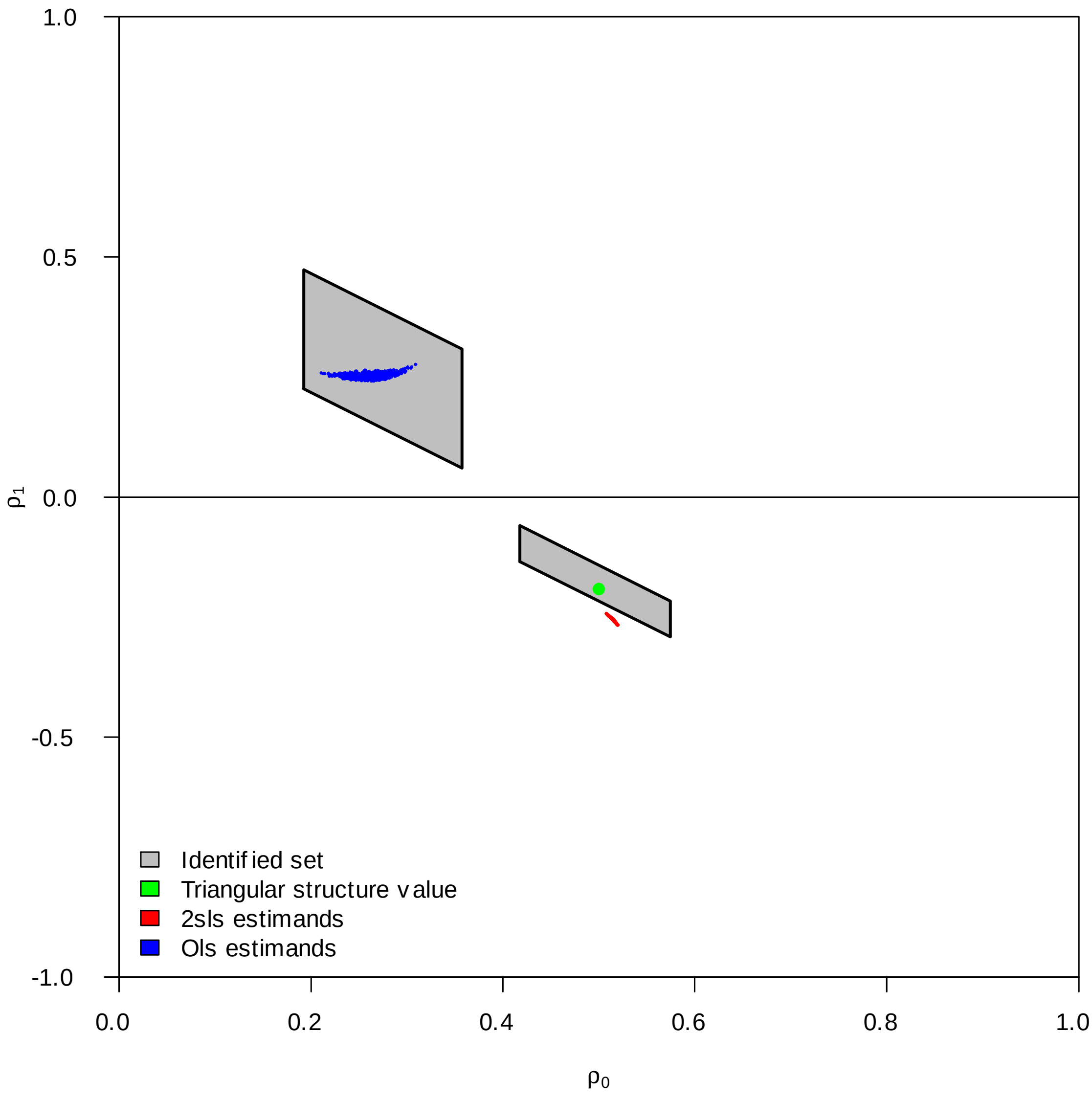

Figure 6: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = +0.7$, $\delta = 0.3$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

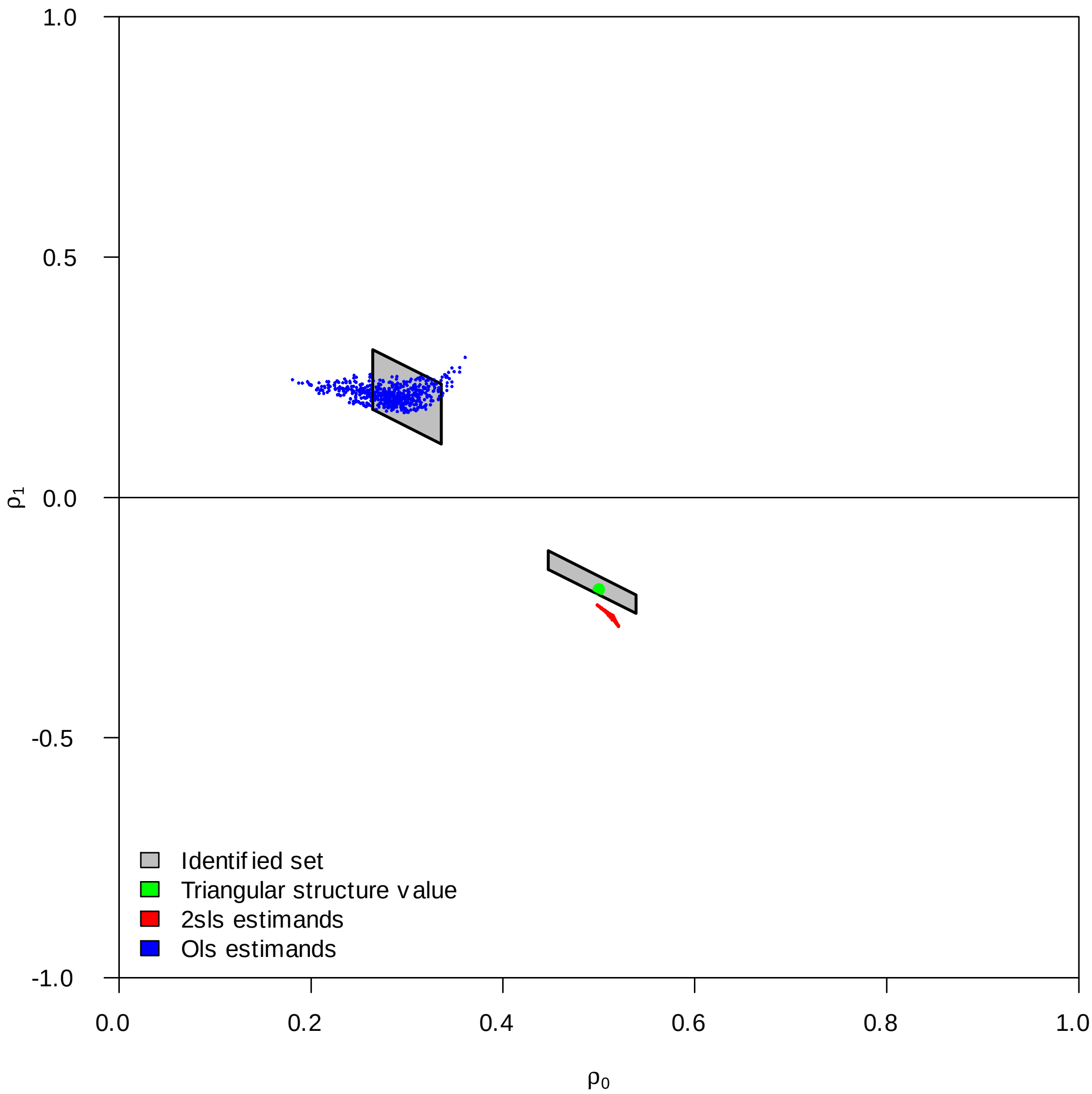

Figure 7: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = +0.7$, $\delta = 0.6$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

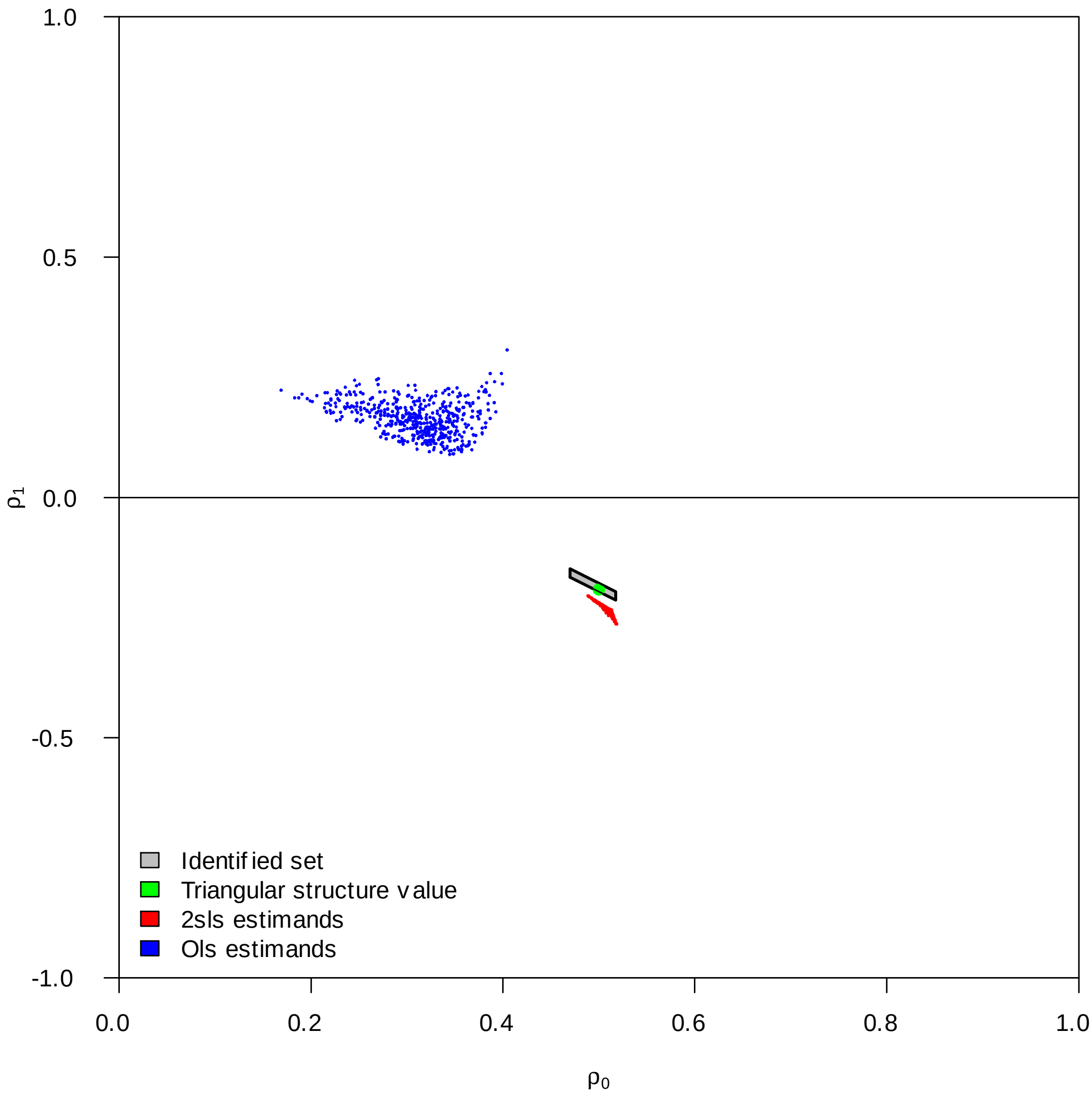

Figure 8: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = +0.7$, $\delta = 0.9$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

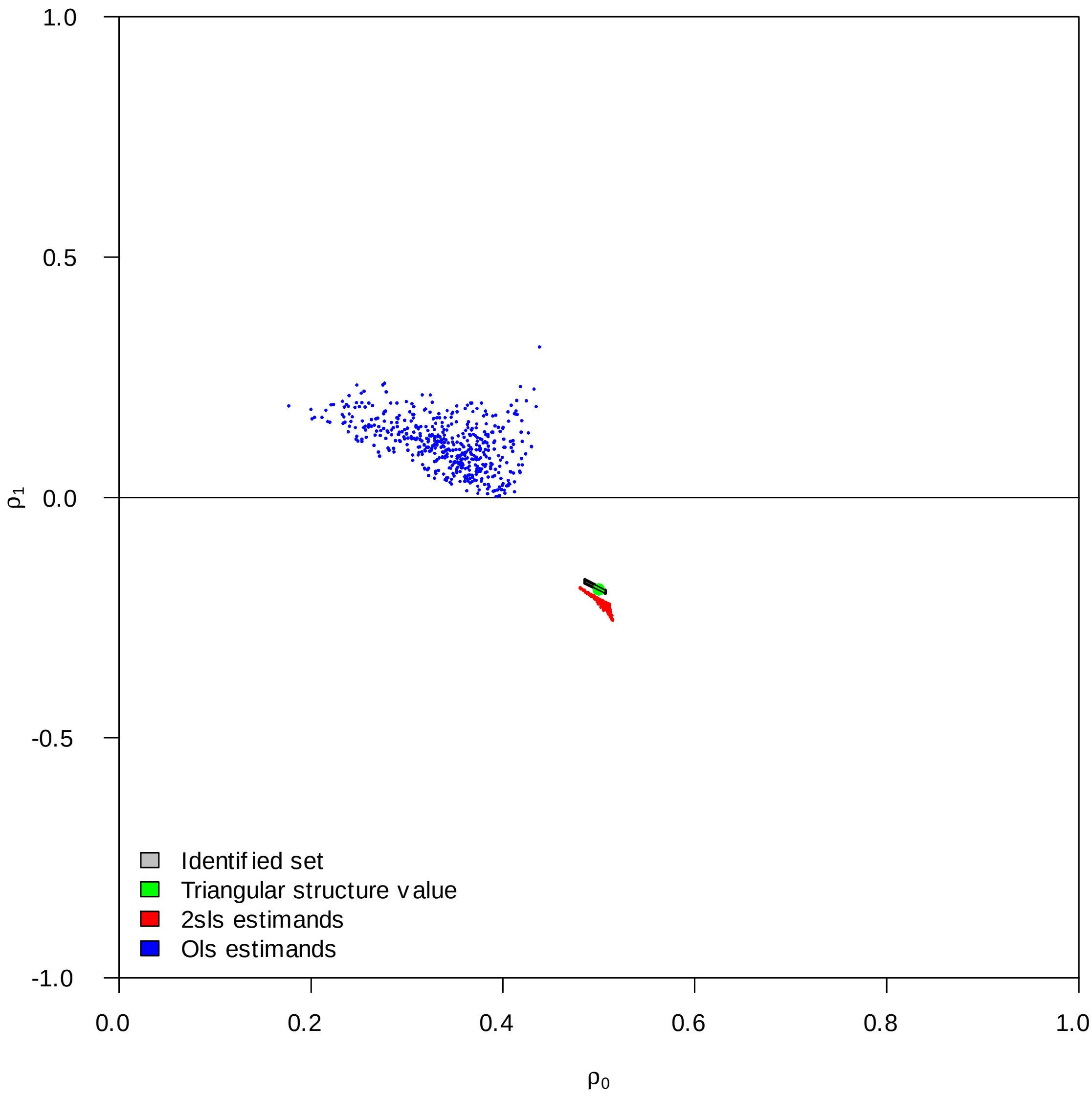

Figure 9: Identified set for (ρ_0, ρ_1) when probability distributions of (Y_1, Y_2) conditional on $Z = z \in \{-1, 0, 1\}$ are generated by the triangular Gaussian structure with $\alpha_\Delta = \gamma = 0$, $\beta_\Delta = -1$, $r = +0.7$, $\delta = 1.2$. 2SLS and OLS estimands are plotted for 500 random distributions of Z.

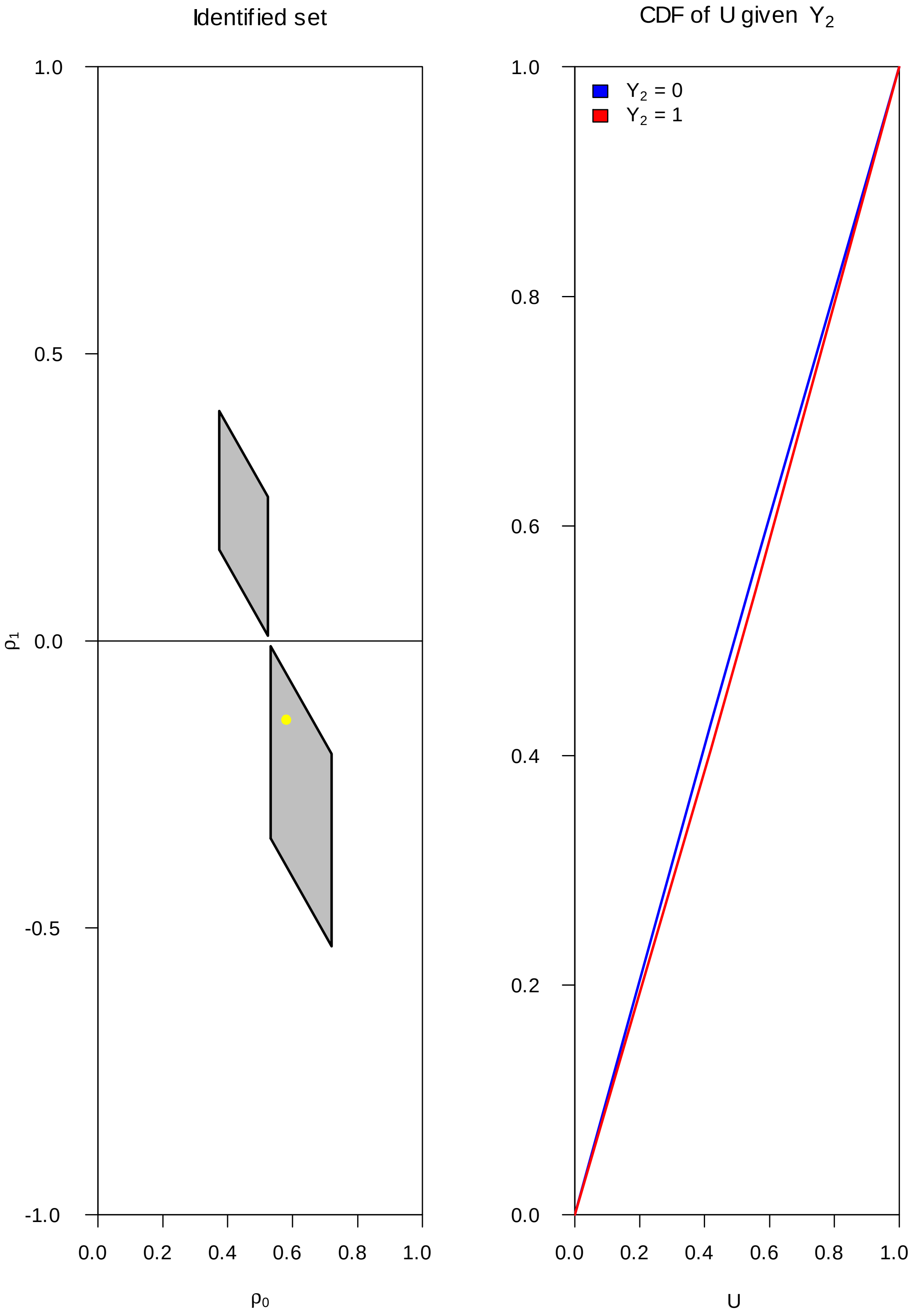

Figure 10: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument. This Figure shows the value of the 2SLS estimate.

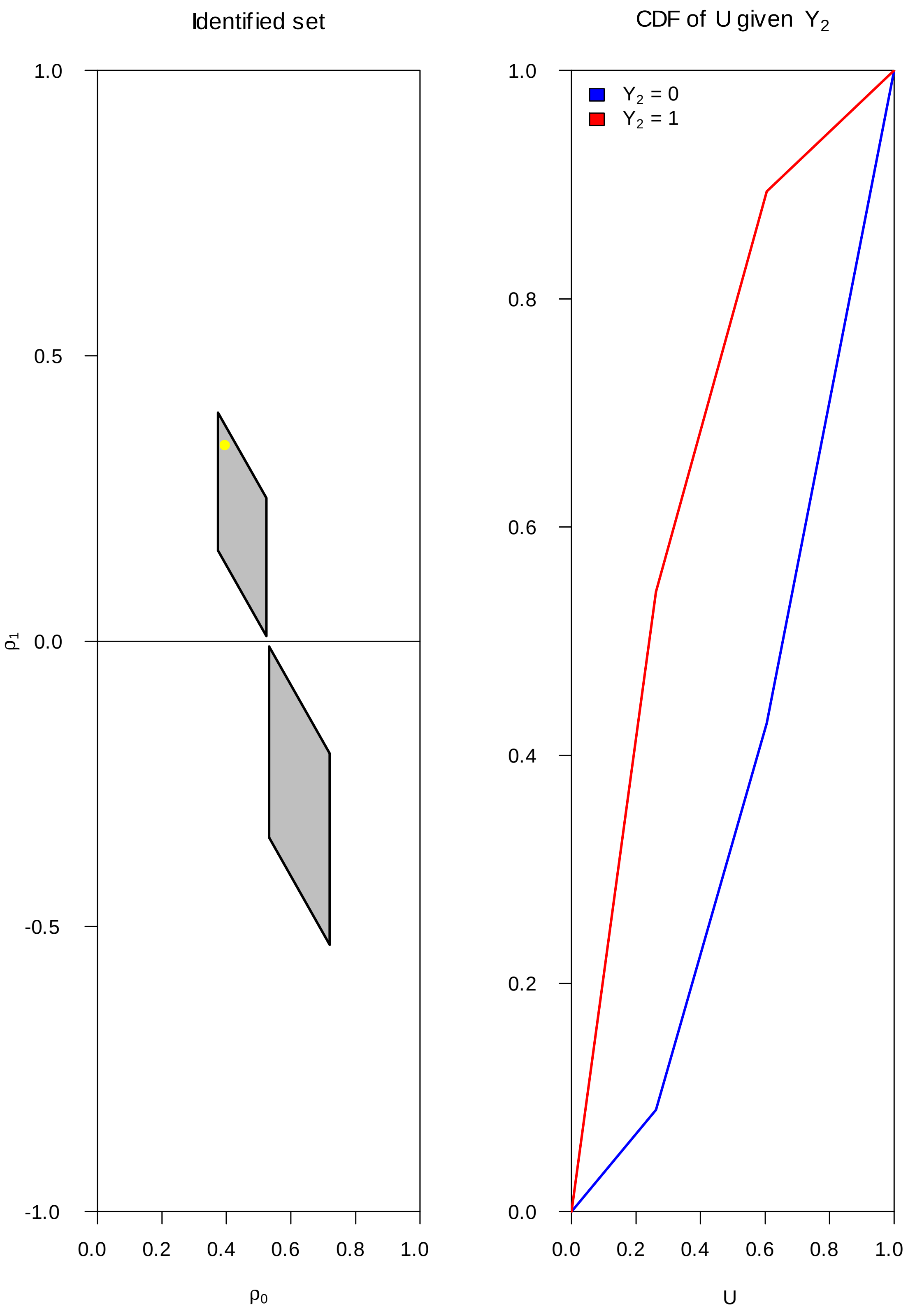

Figure 11: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

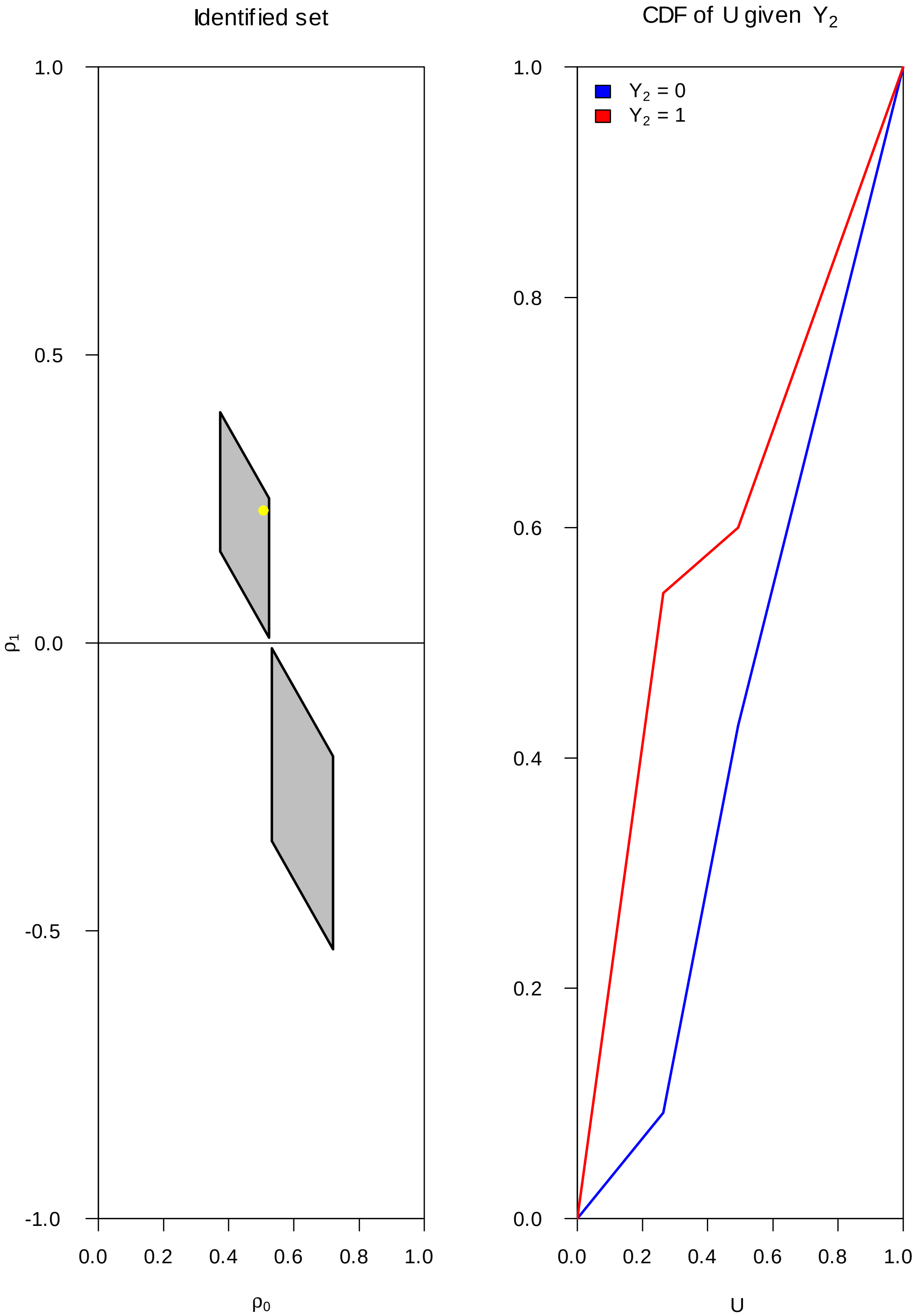

Figure 12: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

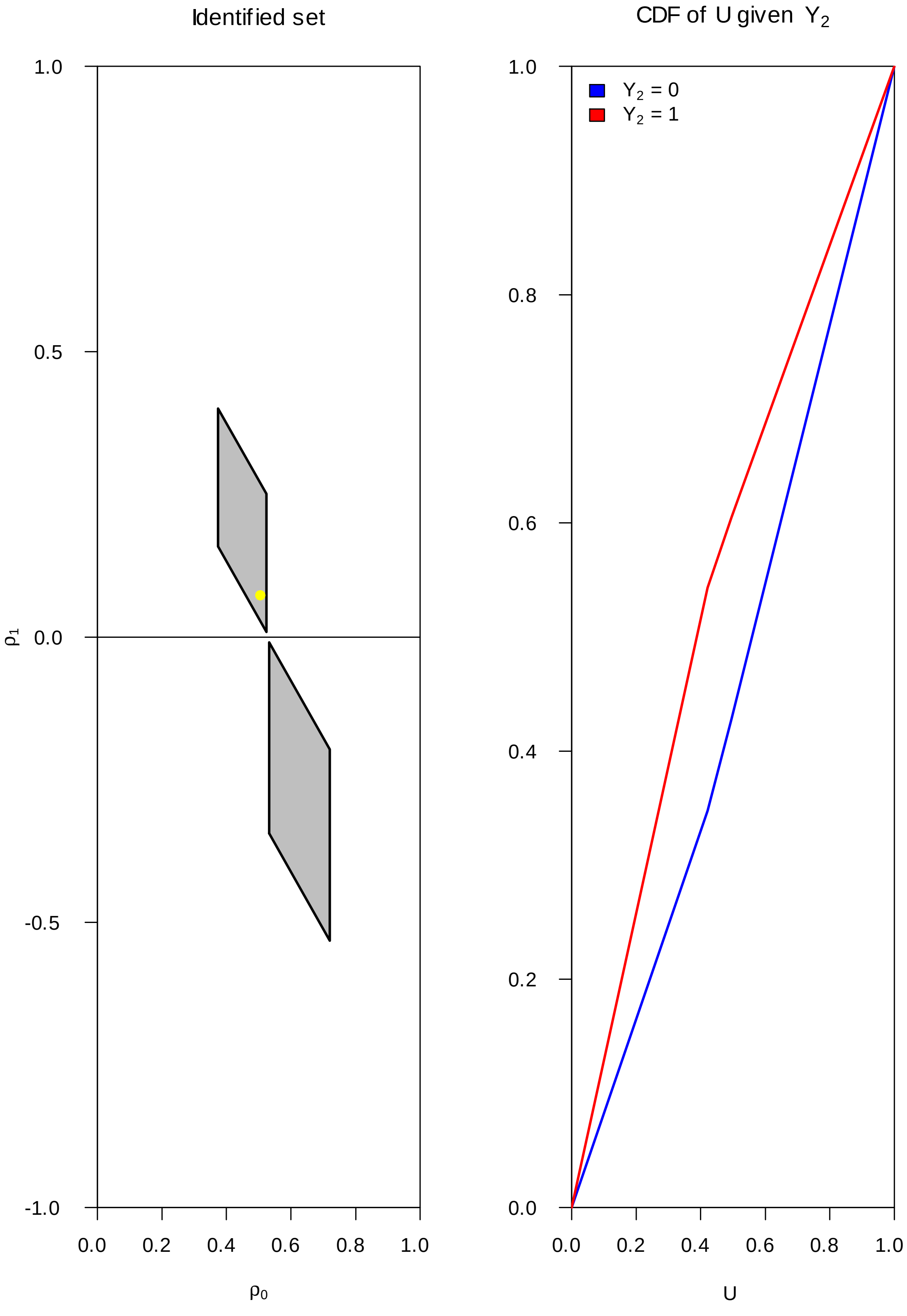

Figure 13: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

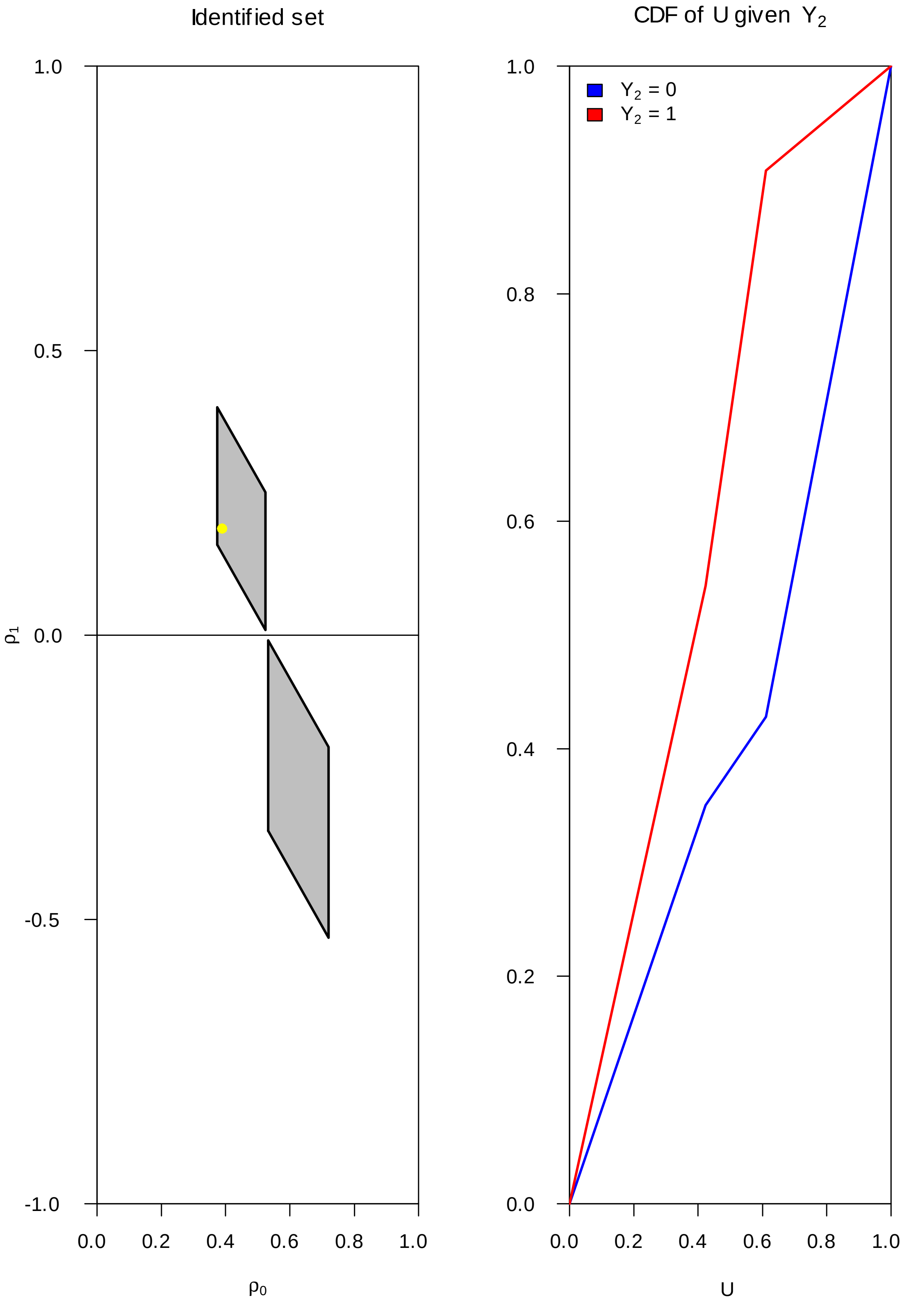

Figure 14: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

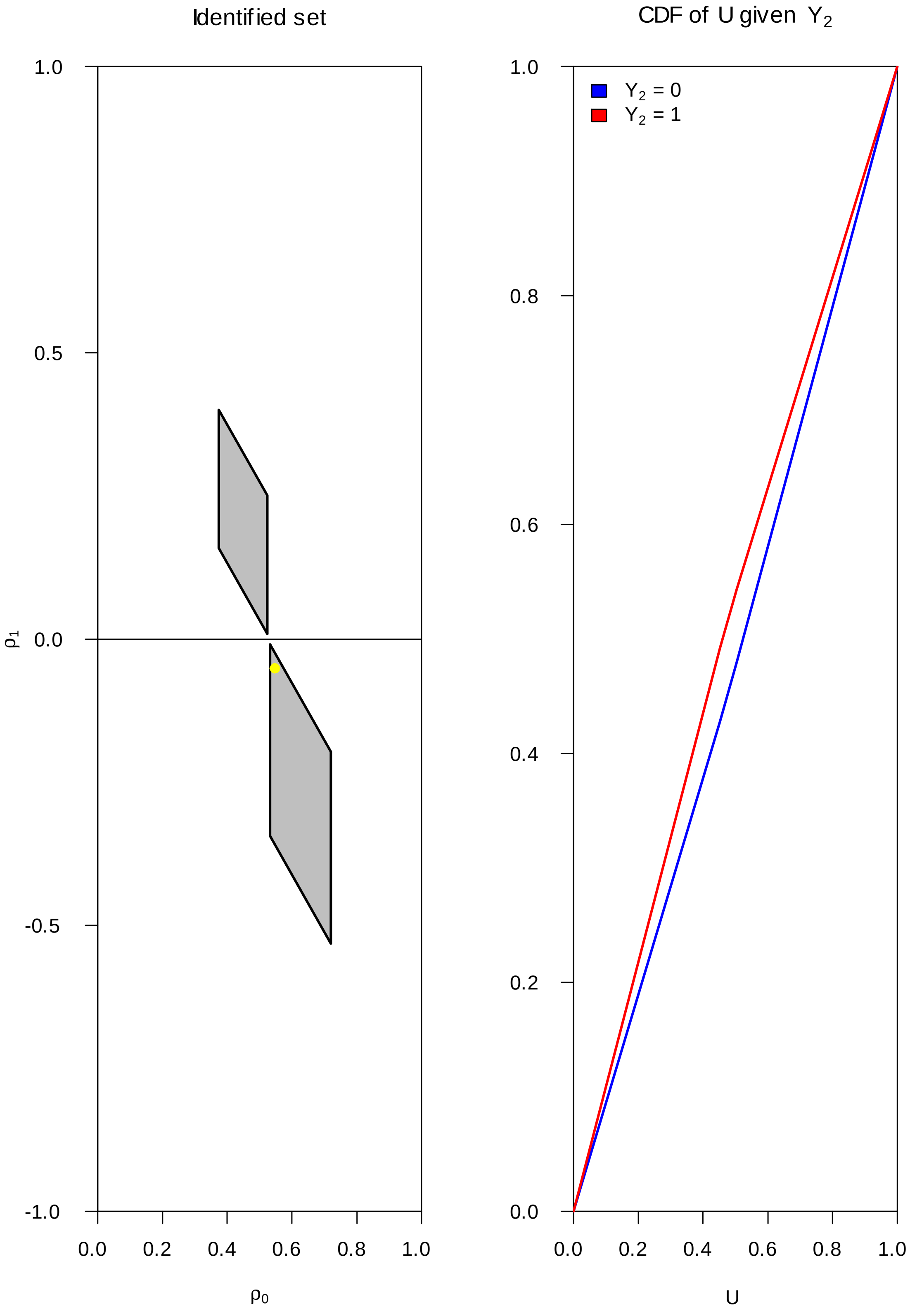

Figure 15: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

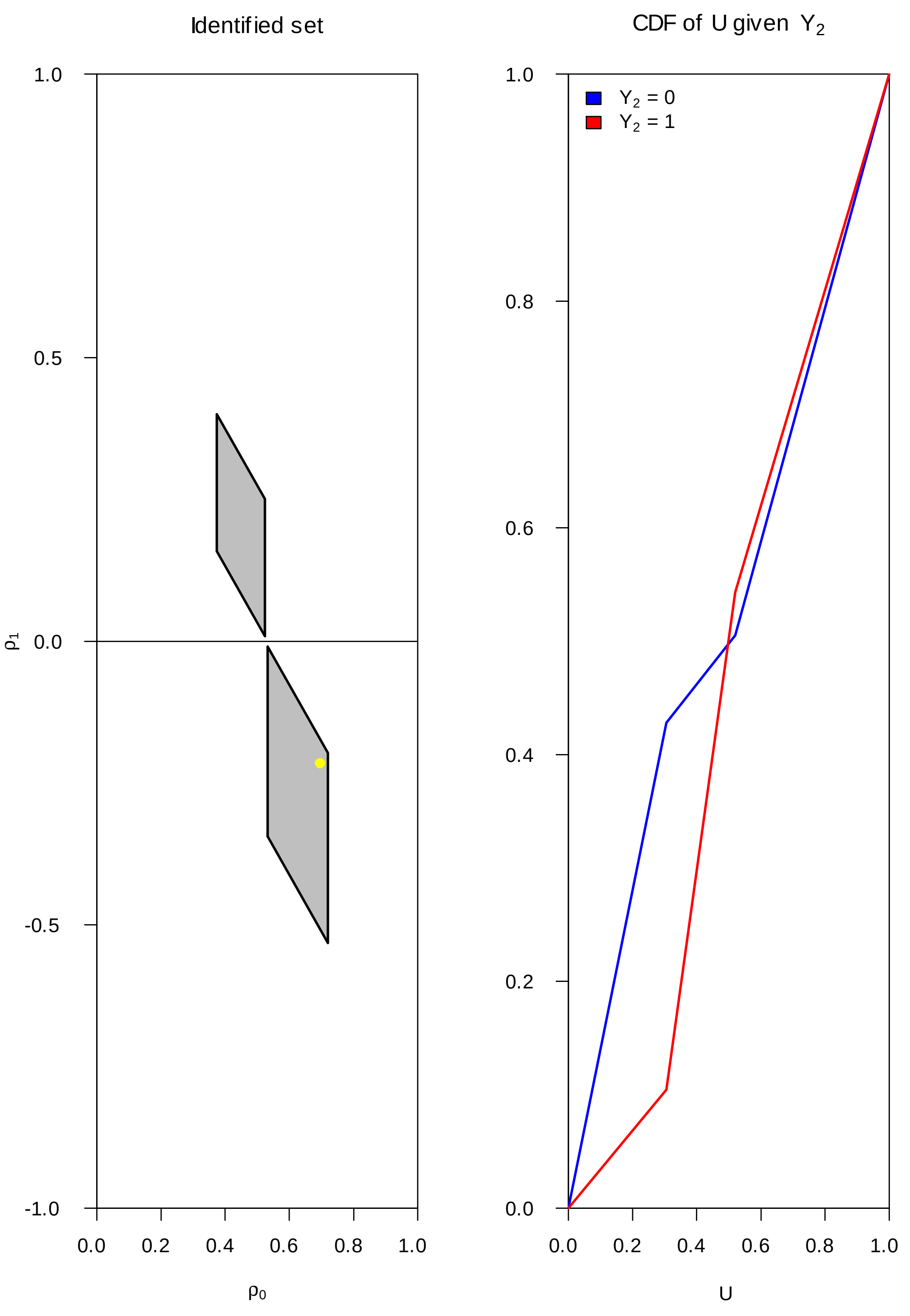

Figure 16: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

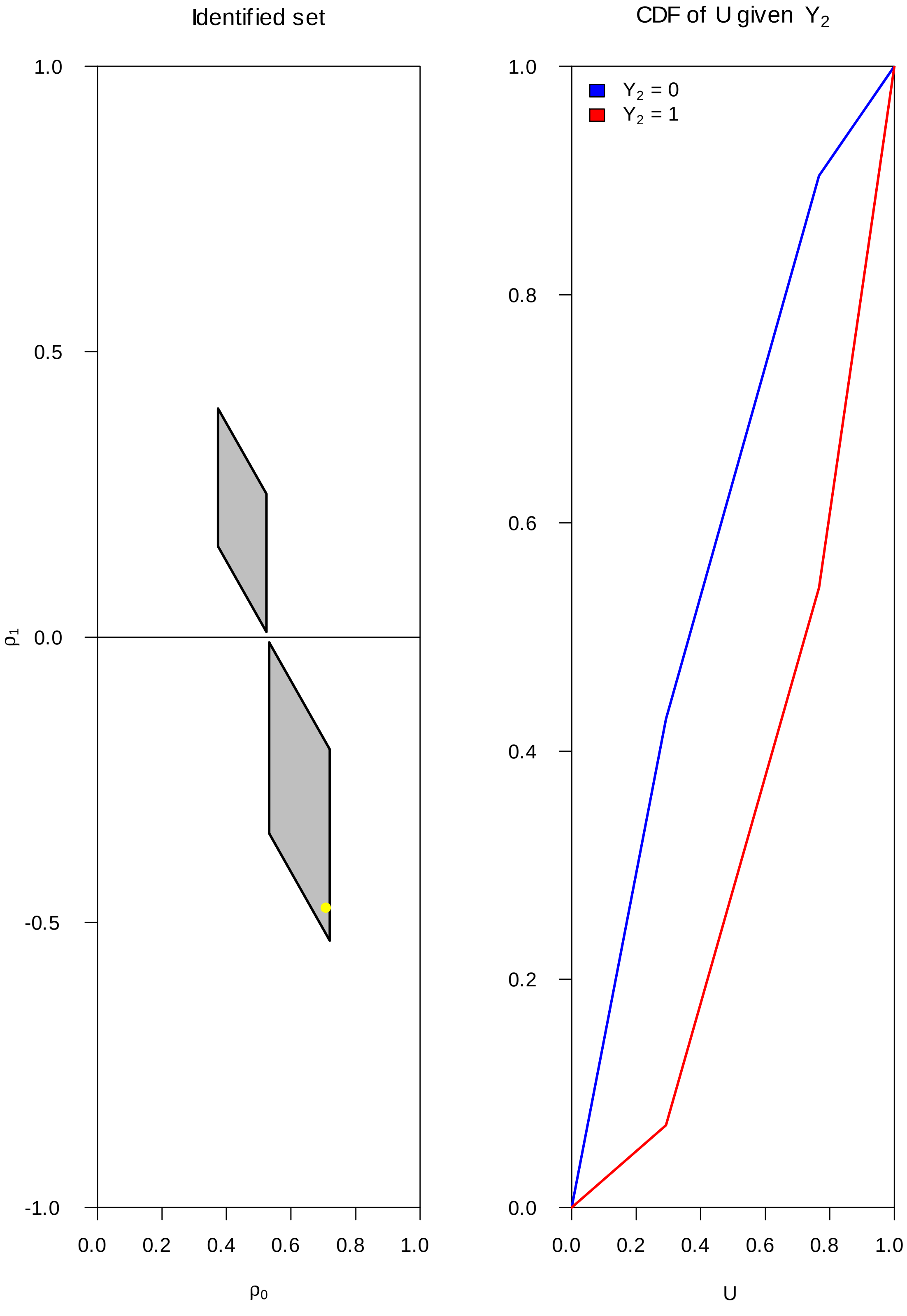

Figure 17: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.

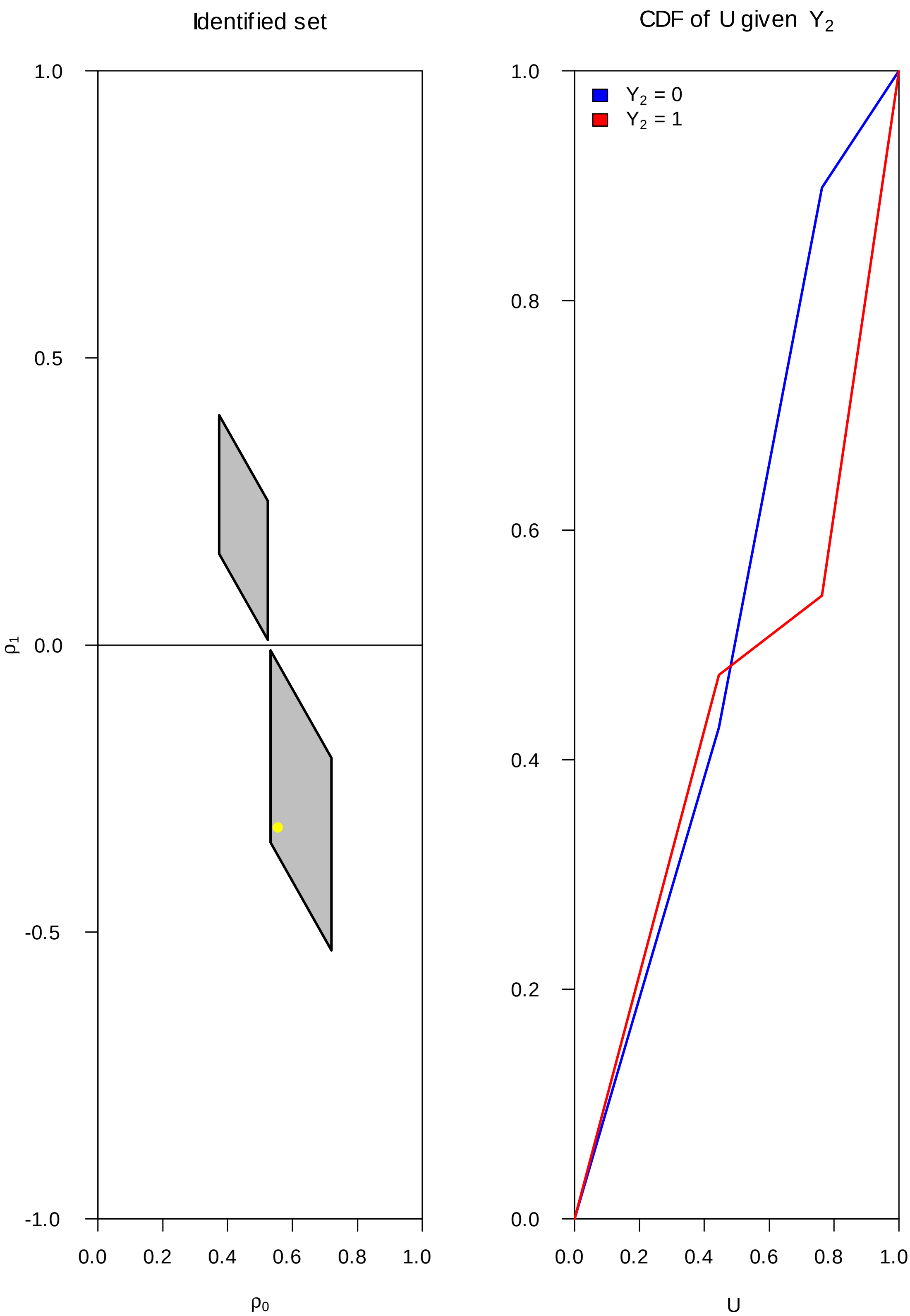

Figure 18: Conditional distribution functions of U given Y_2 (righthand pane) associated with the highlighted value of (ρ_0, ρ_1) in the estimated identified set (left hand pane) obtained with the Angrist and Evans (1998) data and the same-sex instrument.