Simulation-based robust IV inference for lifetime data

Anand Acharya[1] Lynda Khalaf [1]
Marcel Voia [1] Myra Yazbeck [2] David Wensley [3]

[1]Department of Economics
Carleton University

[2]Department of Economics
University of Ottawa

[3]Department of Pediatrics
University of British Columbia

June 9, 2017

Research question, model and complications

- **Research Question** $\Rightarrow$ What is the relationship between a patient's length of stay in the pediatric intensive care unit and their illness severity score at the time of admission.
- **Duration Model** $\Rightarrow$ Accelerated failure time (AFT).
- **Complications** $\Rightarrow$ (i) Unmeasured confounding or endogeneity arising from an omitted variable (unobserved heterogeneity or frailty). (ii) Censoring.
- **Methods** $\Rightarrow$ Robust instrumental variables (IV): the generalized Anderson-Rubin (GAR) statistic and the generalized Andrews-Marmer (GAM) statistic.

Accelerated life model

Underlying assumption is covariates "accelerate" or "decelerate" observed time, by a constant factor, $exp(Y\beta + X_1\delta)$. Expressed as a transformation model:

$$y = \delta_\iota + Y\beta + X_1\delta + \sigma\epsilon. \quad (1)$$

- $y \equiv ln(t)$: transformed possibly right-censored $(n \times 1)$ durations,
- Y : confounded observed $(n \times 1)$ risk scores,
- X_1 : observed $(n \times k_1)$ covariates,
- ϵ : unobserved $(n \times 1)$ random disturbance.

Also observe other $(n \times 1)$ instrumental variables X_2.

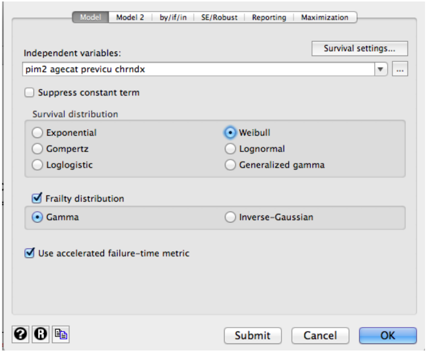
Parametric survival models

- $Lognormal(exp(\delta_\iota), \sigma^2) \rightarrow \epsilon \overset{iid}{\sim} Normal(0,1)$,
- $Loglogistic(exp(\delta_\iota), \sigma) \rightarrow \epsilon \overset{iid}{\sim} Logistic(0,1)$,
- $Weibull(exp(\delta_\iota), \frac{1}{\sigma}) \rightarrow \epsilon \overset{iid}{\sim} Gumbel(0,1)$

where the *Lognormal* location, *Loglogistic* location, and *Weibull* scale parameters are respectively captured in the transformed regression intercept, δ_ι.

```
streg [varlist] [if] [in] [, options]
```

Statistics > Survival analysis > Regression models > Parametric survival models




```
streg PRISM age_cat chrndx previcu, dist(weibull)
```

Assumptions

- Assumption **A 2**: X_1, X_2 predetermined, or
- Assumption **A 3**: X_2, ϵ pairwise stochastically independent.
- Assumption **A 4**: (X_1, ϵ) independently distributed.

- Assumption **D 1**: ϵ distribution unspecified.
- Assumption **D 2,3,4**: $\epsilon \overset{iid}{\sim} Normal(0,1)$, Logistic(0,1) or Gumbel(0,1).

- Assumption **C 3**: $t^* = min(\tau, t)$ and d is the censoring indicator.

Weak Instruments and Identification Robustness

- Explicitly make no assumptions on the data generating process that links Y and X_2 or on the functional form of the first stage regression
- Anderson and Rubin (1949) proposed inverting a least squares test that assesses the exclusion of the instruments in an auxiliary regression.
- *auxiliary* (least squares) *regression*

$$y - Y\beta_o = X_{1\iota}\lambda + X_2\gamma + \omega, \tag{2}$$

where ω is an $(n \times 1)$ random disturbance and $X_{1\iota} = [\iota, X_1]$.

Least Squares Statistic

- Generalize Anderson and Rubin (1949) test statistic for $H_o : \beta = \beta_o \Rightarrow \gamma = 0$:

$$GAR(\beta_o,) = \frac{(y - Y\beta_o)'(M_1 - M)(y - Y\beta_o)/k_2}{(y - Y\beta_o)'M(y - Y\beta_o)/(n-k)}, \qquad (3)$$

where $M = I - X(X'X)^{-1}X'$, in which $X = [X_{1\iota}, X_2]$ and $M_1 = I - X_{1\iota}(X'_{1\iota}X_{1\iota})^{-1}X'_{1\iota}$.

- Pivotal statistic $\Rightarrow$ Exact null distribution:

$$\overline{GAR}(\beta_o) = \frac{\epsilon'(M_1 - M)\epsilon/k_2}{\epsilon' M\epsilon/(n-k)}, \Rightarrow gar_{calc}(\alpha), \qquad (4)$$

Robust inference

To construct a confidence set on β_o, we invert[1] a generalized Anderson-Rubin (GAR) statistic derived from an auxiliary regression:

$$C_\beta(\alpha) = \{\beta_o : GAR(\beta_o) < gar_{calc}(\alpha)\}, \quad (5)$$

Solution permits sets that are *closed, open, empty, or the union of two or more disjoint intervals.*[2]

[1] Dufour & Taamouti(2005)

[2] Dufour(1997)

$$C_\beta(\alpha) = \{\beta_o : \beta_o' A \beta_o + b' \beta_o + c \le 0\},$$

- $(n \times 1)$ vector $\mathbf{u}_j$ is drawn from the uniform [0,1]
- jth realization of the GAR statistic
- Repeat for j=*1..J*.
- Construct the simulated exact null distribution.
- Appropriate α-level cut off $\rightarrow$ confidence set construction.

```
M=I(n)-x*luinv(x'x)*x'
M1=I(n)-x1*luinv(x1'x1)*x1'
M12=M1*x2
P=M12*luinv(M12'M12)*M12'

T=J(1000,1,.)
for (i=1; i<=1000; i++) {
u = uniform(n,1)
mu=ln(u:/(1:-u)) /* mu_g=-ln(-ln(u)) */
GAR=((mu'(P*mu))/(mu'(M*mu)))*(n-k)
T[i]=GAR
}

Tc=sort(T,1)
Tcalc=Tc[950] /* Tcalc=invFtail(k2,n-k,.05) */

G=Tcalc/(n-k)
C=P-(G*M)
a=Y'C*Y
b=-2*y'C*Y
c=y'C*y
det=(b*b) - (4*a*c)
d=sqrt(det)
betaL=(-b-d)/(2*a)
betaU=(-b+d)/(2*a)
```

Aligned linear rank statistic.[3]

- Generalize Andrews and Marmer (2008) test statistic for $H_o : \beta = \beta_o \Rightarrow \gamma = 0$:

$$rank(y - Y\beta_o - x_1\hat{\delta}(\beta_o)) = x_2\gamma + \omega, \tag{6}$$

- Test statistic:

$$GAM(\beta_o) = c(i)'\,(p_2)\,c(i), \tag{7}$$

where: $p_2 = x_2(x_2'x_2)^{-1}x_2'$

- c is a score vector of: $(i) = rank(y - Y\beta_o - x_1\hat{\delta})$.

[3] Andrews and Marmer (2008)

Rank scores.

- Rank scores are derived to be efficient for certain distributional specifications, F_o.
- However, they are robust to misspecification.[4].
- The score vector satisfy a non-decreasing and non-constant condition, $c^{(i)} \leq ... \leq c^{(n)}$ and $c^{(i)} \neq c^{(n)}$, where (i) is the rank label of the associated aligned residual order statistic.
- Two related and asymptotically equivalent scores are the quantile F_o scores and the expected value F_o scores.

[4] Chernoff and Savage (1958)

Rank scores: Quantile and expected value.[5]

- Quantile F_o scores:

$$c^{(i)} = F_o^{-1}\left(\frac{(i)}{(n+1)}\right). \tag{8}$$

- Expected value F_o scores:

$$c^{*(i)} = E_{F_o}[V^{(i)}], \tag{9}$$

where $V^{(i)}$ is the ith order statistic in a random sample of size n and (i) is the rank label of the associated aligned residual order statistic.

[5] Randles and Wolfe (1979)

Quantile scores.

- Quantile scores use the rank label to reconstruct the variate values from the quantile function of a presumed distribution.
- Normal quantile function of VanderWaerden (1953):

$$c^{(i)} = \Phi^{-1}((i)^{*}). \tag{10}$$

- Logistic:

$$c^{(i)} = ln(\frac{(i)^{*}}{1-(i)^{*}}) \tag{11}$$

- Gumbel:

$$c^{(i)} = -ln(-ln((i)^{*})). \tag{12}$$

Where $(i)^{*} = \left(\frac{(i)}{(n+1)}\right)$

Mata code: Quantile scores

```
for (i=1; i<=1000; i++) {
    u = uniform(n,1)
    R=mm_ranks(u,1,0)
    O=R,u,x2d,d
    Os=sort(O,1)
    X2=Os[,3]
    p2=X2*luinv(X2'X2)*X2'
    Rs=Os[,1]
    ds=Os[,4]
    Rp=Rs:/(n+1)

    RQ_n = invnormal(Rp)
    RQ_l = ln(Rp:/(1:-Rp))
    RQ_g = -ln(-ln(Rp))
    RQ_e = -ln(1:-Rp)

    RAR_n=RQ_n'p2*RQ_n
    RAR_l=RQ_l'p2*RQ_l
    RAR_g=RQ_g'p2*RQ_g
    RAR_e=RQ_e'p2*RQ_e

    Tn[i]=RAR_n
    Tl[i]=RAR_l
    Tg[i]=RAR_g
    Te[i]=RAR_e
}
```

```
for (i=1; i<=nb; i++) {
E=M1d*(y-Y*gridBeta0[i])
    R=mm_ranks(E,1,0)
    O=R,E,x2d,d
    Os=sort(O,1)
    X2=Os[,3]
    p2=X2*luinv(X2'X2)*X2'
    Rs=Os[,1]
    ds=Os[,4]
    Rp=Rs:/(n+1)

    RQ_n = invnormal(Rp)
    RQ_l = ln(Rp:/(1:-Rp))
    RQ_g = -ln(-ln(Rp))
    RQ_e = -ln(1:-Rp)

    RAR_n=RQ_n'p2*RQ_n
    RAR_l=RQ_l'p2*RQ_l
    RAR_g=RQ_g'p2*RQ_g
    RAR_e=RQ_e'p2*RQ_e

Rx[i,1]=gridBeta0[i]
Rx[i,2]=RAR_n
Rx[i,3]=RAR_l
Rx[i,4]=RAR_g
Rx[i,5]=RAR_e
```

Expected value scores.

- Well know classical expected value scores:
- Wilcoxon (1945), where the expected value of the order statistic is derived from sampling the logistic distribution, giving:

$$c^{*(i)} = \frac{2(i)}{(n+1)} - 1.$$

```
RE_w=((2:*Rs):/(n+1)):-1
```

- Savage (1956), where the expected value of the order statistic is derived from sampling the exponential distribution, giving:

$$c^{*(i)} = \frac{1}{n} + \frac{1}{(n-1)} + \ldots + \frac{1}{(n-(i)+1)} - 1$$

```
RE_s=runningsum(1:/(n:-Rs:+1)):-1
```

Right censoring.

- We assume a right censoring scheme in which the censoring indicator, d is independently distributed.
- Where observed time is now, $t^* = min(\tau, t)$ in which τ is the censored time.
- Utilize the framework of Prentice (1978) to adjust the rank scores for right censoring.
- Index each censored observation within any adjacent non-censored pair by m.
- All censored observations within the same non-censored interval receive the same score.
- Conceptually, all censored observations now contribute to the rank vector probability via their survivor function.
- May only be applied to expected value scores.

Right censoring.

Utilizing the above framework, the expected value rank scores[6] are:

- Wilcoxon (1945)

$$c^{(i)} = 1 - 2\prod_{j=1}^{i}\frac{n_j}{n_j+1}, \quad c_{m_i}^{(i)} = 1 - \prod_{j=1}^{i}\frac{n_j}{n_j+1}.$$

- Savage (1956)

$$c^{(i)} = \sum_{j=1}^{i} n_j^{-1} - 1, \quad c_{m_i}^{(i)} = \sum_{j=1}^{i} n_j^{-1},$$

where n_j denotes the number of individuals at risk commencing period $t_{(j)}$.

[6] Kalbfleisch and Prentice (2002) Chapter 7

Mata code

Wilcoxon

```
R_set=J(n,1,.)
    R_set[1]=n
        for (j=2; j<=n; j++)
        {
        if (ds[j]==0)    R_set[j]=n+1-Rs[j]
        else             R_set[j]=R_set[j-1]
        }
                         n_j=R_set:/(R_set:+1)
    F=J(n,1,.)
    F[1]=n_j[1]
        for (j=2; j<=n; j++)
        {
        if (ds[j]==0)    F[j]=n_j[j]*F[j-1]
        else             F[j]=F[j-1]
        }
    c=J(n,1,.)

        for (j=1; j<=n; j++)
        {
        if (ds[j]==0)    c[j]=1-2*F[j]
        else             c[j]=1-F[j]
        }
RAR_w_C=c'p2*c
Twc[i]=RAR_w_C
```

Savage

```
                         iR_set=1:/R_set
    Fs=J(n,1,.)
    Fs[1]=iR_set[1]

        for (j=2; j<=n; j++)
        {
        if (ds[j]==0)    Fs[j]=iR_set[j]+Fs[j-1]
        else             Fs[j]=Fs[j-1]
        }

    cs=J(n,1,.)

        for (j=1; j<=n; j++)
        {
        if (ds[j]==0)    cs[j]=Fs[j]-1
        else             cs[j]=Fs[j]
        }

RAR_s_C=cs'p2*cs
Tsc[i]=RAR_s_C
```

Simulation

Empirically relevant simulation design adopts the data generating process:

$$y = Y\beta + X_1\delta + \epsilon, \quad Y = h(X_1\pi_1 + X_2\pi_2 + \sqrt{1-\rho^2}\mu + \rho\epsilon),$$

Size control is achieved in all specifications. Power is increasing in:

- Instrument strength.
- Instrument balance.
- Effect size (clinically relevant difference).
- Sample size.

Clinical research question

Research Question $\Rightarrow$ What is the relationship between a patient's length of stay (LoS) in the pediatric intensive care unit (PICU) and their illness severity score at the time of admission?

- **Outcome** $\Rightarrow$ Pediatric intensive care unit length of stay (LoS_i) measured in hours.
- **Exposure** $\Rightarrow$ Illness severity index as a marker of the exposure, as measured by either $PIM2_i$ and $PRISMIII_i$.
- **Data** $\Rightarrow$ Prospectively collected observational data set. Five centres and $i = 1...10,044$ patients over a two year period representing 1,184,726 PICU hours.

Complications

Primary complication ⇒ Unmeasured factors may affect both exposure (illness severity) and outcome (LoS). Since randomized control study design may not be feasible, the use of instrumental variables provides one possible solution to this problem.

Secondary complication ⇒

- Long stay (>10 days) (1,078/10,044) 12 % of sample used (663,368/1,184,726hrs) 56 % of PICU hours.
- Death (354/10,044) 3.5 % of sample used (122,766/1,184,726hrs) 10.4 % of PICU hours.
- Trauma (658/10,044) 6.6 % of sample used (69,869/1,184,726hrs) 5.9 % of PICU hours.

Pediatric illness severity scores and risk adjustment

- Pediatric Index of Mortality (*PIM*2)[7] and Pediatric Risk of Mortality (*PRISMIII*)[8]
- Derived from a patient's probability of mortality, but primarily used as measure of illness severity.
- Employed in risk-adjusting outcomes and stratifying patients.
- Imperfect signal on patient's "type".
- Do not account for individual specific effects.

[7] Slater et al (2004)
[8] Pollack et al (1996)

Clinical Model and Trauma

$$ln(LoS_i) = \delta_\iota + \beta PIM2_i + \delta_A Agecat_i + \delta_C Chrndx_i + \delta_P Previcu_i + \sigma\epsilon_i.$$

- Where the illness severity index $PIM2_i$ is confounded. Instrumental variables are a possible solution $\Longrightarrow Trauma_i$
- The selection of the instrument was based on the intuition that a patient that suffered a trauma was *as good as randomly assigned*, in the context of the clinical model.
- The otherwise unobserved heterogenous types would be equally as likely to suffer a trauma.

Results: 95% Confidence Sets for β.

		PIM2 Continuous Bimodal	*PRISM* Categorical Point-mass at 0
Log-logistic	AFT	(.294, .321)	(.096, .104)
	Gamma-frailty	(.287, .314)	(.095, .103)

		PIM2 Continuous Bimodal	*PRISM* Categorical Point-mass at 0
GAR	Least-squares	(.070, .193)	(.104, .305)
GAM	Quantile	(.065, .175)	(.115, .415)
	Wilcoxon	(.040, .160)	(.180, .440)
Censored	Wilcoxon	(.070, .240)	(.190, .750)

Length of stay and illness severity

- Trauma is an informative instrument.
- Illness severity, as measured by *PIM*2 or *PRISMIII*, appears to be confounded.
- The difference in effect size has both clinical and policy relevance.
- The robust procedure exploits data otherwise often ignored from analysis: (i) Trauma (ii) Mortality and (iii) Long stay.

Conclusion

- Clinically relevant question with useful policy implications.
- Proposed a novel method of robust inference.
- Extended the identification robust instrumental variables approach to duration analysis.
- Unmeasured factors may affect both intervention and outcome. In situations where randomized control study design may not be feasible, the use of robust instrumental variables provides one possible solution to this problem.

Anderson, T. W. & Rubin, H. (1949).
Estimation of the parameters of a single equation in a complete system of stochastic equations.
The Annals of Mathematical Statistics **20(1)**, 46–63.

Andrews, D. W. K. & Marmer, V. (2008).
Exactly distribution free inference in instrumental variables regression with possibly weak instruments.
Journal of Econometrics **142**, 183–200.

Chernoff, H. & Savage, I. R. (1958).
Asymptotic normality and efficiency of certain non-parametric test statistics.
The Annals of Mathematical Statistics **29**, 972–994.

Dufour, J. M. (1997).
Some impossibility theorems in econometrics with applications to structural and dynamic models.
Econometrica **65**, 1365–1387.

DUFOUR, J. M. & TAAMOUTI, M. (2005).
Projection-based statistical inference in linear structural models with possibly weak instruments.
Econometrica **4**, 1351–1365.

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002).
The Statistical Analysis of Failure Time Data.
John Wiley & Sons.

POLLACK, M. M., PATEL, K. & RUTTIMANN, U. E. (1996).
PRISM III: An updated Pediatric Risk of Mortality score.
Critical Care Medicine **24**, 743–752.

SLATER, A., SHAN, F. & PEARSON, G. (2003).
PIM2: A revised version of the paediatric index of mortality.
Intensive Care Medicine **29**, 278–285.

PRENTICE, R. L. (1978).
Linear rank tests with right censored data.
Biometrika **65**, 167–179.

RANDLES, R. H. & WOLFE, D. A. (1979).
Introduction to the Theory of Nonparametric Statistics.
John Wiley & Sons.

SAVAGE, I. R. (1956).
Contributions to the theory of rank order statistics – the two-sample case.
Ann. Math. Statist. **27**, 590–615.