PLSC 504 – Fall 2020
Models for Ordinal Outcomes And Event Counts

September 9, 2020

Ordinal Data

Ordinal data are:

- Discrete: $Y \in \{1, 2, ...\}$
- *Grouped Continuous* Data
- *Assessed Ordered* Data

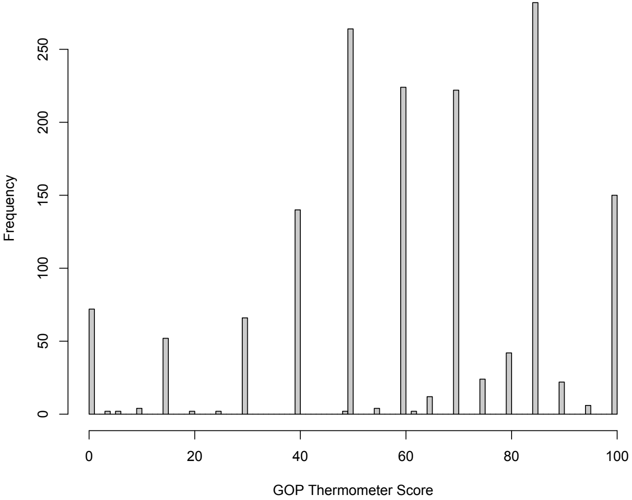
In general:

- Some things can be ordered, but shouldn't be
- Some things are ordered in some circumstances but not others
- Orderings can differ across applications

Ordinal vs. Continuous Response Models

"I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 and 100 degrees mean that you feel favorably and warm toward the person; ratings between 0 and 50 degrees mean that you don't feel favorably toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person."

GOP Thermometer Scores (1988)



$$Y_i^* = \mu + u_i$$

$$Y_i = j \text{ if } \tau_{j-1} \leq Y_i^* < \tau_j,\ j \in \{1, ...J\}$$

$$\begin{aligned} Y_i &= 1 \text{ if } -\infty \leq Y_i^* < \tau_1 \\ &= 2 \text{ if } \tau_1 \leq Y_i^* < \tau_2 \\ &= 3 \text{ if } \tau_2 \leq Y_i^* < \tau_3 \\ &= 4 \text{ if } \tau_3 \leq Y_i^* < \infty \end{aligned}$$

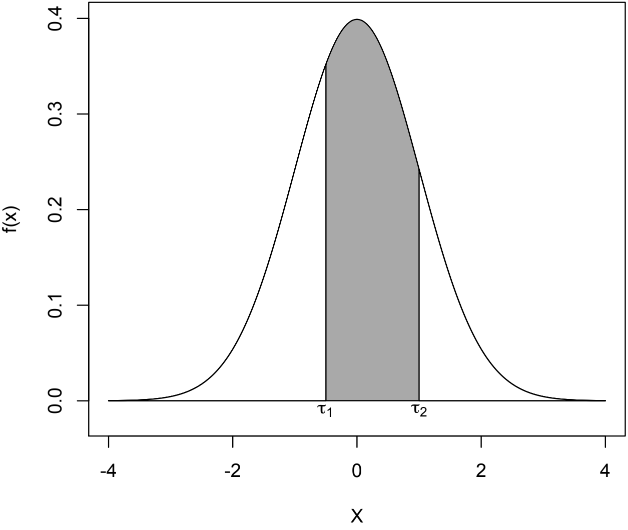
Ordinal Response Models: Probabilities

$$
\begin{aligned}
\Pr(Y_i = j) &= \Pr(\tau_{j-1} \leq Y^* < \tau_j) \\
&= \Pr(\tau_{j-1} \leq \mu_i + u_i < \tau_j)
\end{aligned} \tag{1}
$$

$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}$$

$$
\begin{aligned}
\Pr(Y_i = j | \mathbf{X}, \boldsymbol{\beta}) &= \Pr(\tau_{j-1} \leq Y_i^* < \tau_j | \mathbf{X}) \\
&= \Pr(\tau_{j-1} \leq \mathbf{X}_i \boldsymbol{\beta} + u_i < \tau_j) \\
&= \Pr(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta} \leq u_i < \tau_j - \mathbf{X}_i \boldsymbol{\beta}) \\
&= \int_{-\infty}^{\tau_j - \mathbf{X}_i \boldsymbol{\beta}} f(u_i) du - \int_{-\infty}^{\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta}} f(u_i) du \\
&= F(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})
\end{aligned}
$$

What That Looks Like



Probabilities (here, probit)

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Phi(\tau_1 - \mathbf{X}_i\boldsymbol{\beta}) - 0 \\
\Pr(Y_i = 2) &= \Phi(\tau_2 - \mathbf{X}_i\boldsymbol{\beta}) - \Phi(\tau_1 - \mathbf{X}_i\boldsymbol{\beta}) \\
\Pr(Y_i = 3) &= \Phi(\tau_3 - \mathbf{X}_i\boldsymbol{\beta}) - \Phi(\tau_2 - \mathbf{X}_i\boldsymbol{\beta}) \\
\Pr(Y_i = 4) &= 1 - \Phi(\tau_3 - \mathbf{X}_i\boldsymbol{\beta})
\end{aligned}
$$

Likelihoods

Define:

$$\begin{aligned}\delta_{ij} &= 1 \text{ if } Y_i = j \\ &= 0 \text{ otherwise.}\end{aligned}$$

Likelihood:

$$L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \prod_{i=1}^{N} \prod_{j=1}^{J} [F(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]^{\delta_{ij}}$$

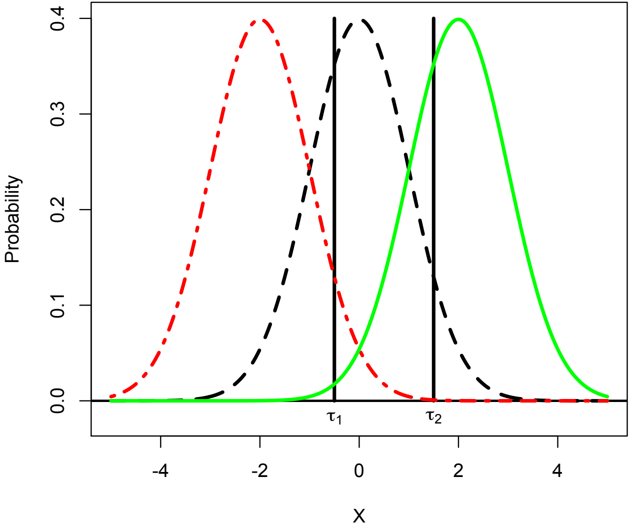
Log-Likelihood, probit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Phi(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Phi(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

Log-Likelihood, logit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Lambda(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Lambda(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

The Intuition



Basic Models: Ordered Logit / Probit

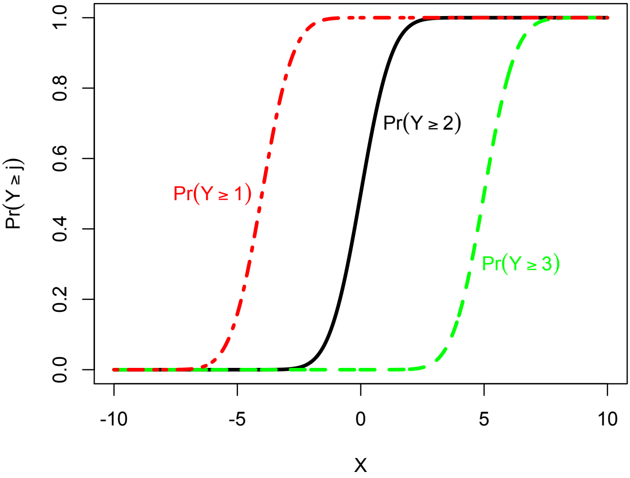
As discussed in PLSC 503 (slides and code):

- Identification
- Estimation / Model Fitting
- Interpretation:
  - Marginal Effects
  - Odds Ratios
  - Predicted Probabilities (including c.i.s and plots)

$$\frac{\partial \Pr(Y_i \geq j)}{\partial X} = \frac{\partial \Pr(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

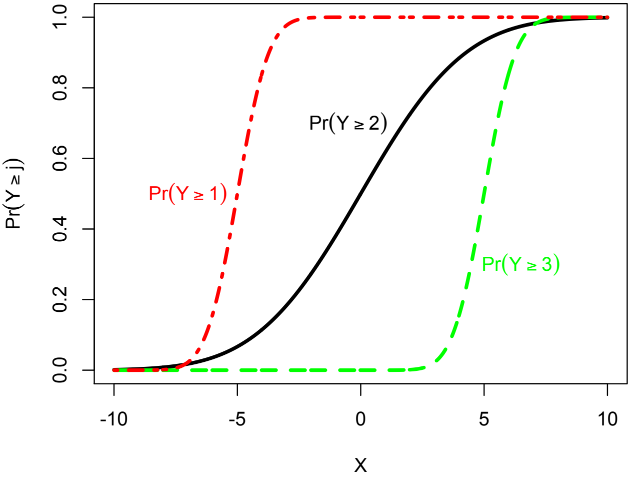
(aka "proportional odds"...)

Parallel Regressions Envisioned



$$\frac{\partial \Pr(Y_i \geq j)}{\partial X} \neq \frac{\partial \Pr(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

Nonparallel Regressions Envisioned



A Generalization

$$\Pr(Y_i = j|\mathbf{X}, \boldsymbol{\beta}) = F(\tau_j - \mathbf{X}_i\boldsymbol{\beta}_j) - F(\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta}_j)$$

- Akin to $J - 1$ binary logits/probits
- Compare using LR/Wald test
- Also Brant (1990)
- Available (canned) in Stata

Other Variants: Heteroscedastic

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln \left[\Phi \left(\frac{\tau_j - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) - \Phi \left(\frac{\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) \right]$$

- See (e.g.) Alvarez and Brehm (1998)

Other Variants: Varying τs

Sanders:

$$
\begin{aligned}
\Pr(Y_i = 1) &= 1 - \Phi\left(\frac{\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right), \\
\Pr(Y_i = 2) &= \Phi\left(\frac{\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right) - \Phi\left(\frac{-\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right), \text{and} \\
\Pr(Y_i = 3) &= \Phi\left(\frac{-\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right).
\end{aligned}
$$

$$
\ln L = \sum_{i=1}^{N}\sum_{j=1}^{J}\delta_{ij}\ \ln\left[\Phi\left(\frac{\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right) - \Phi\left(\frac{-\mathbf{W}_i\eta - \mathbf{X}_i\boldsymbol{\beta}}{\exp(\mathbf{Z}_i\gamma)}\right)\right]
$$

- Maddala (1983); Terza (1985)
- "Cut points" are symmetrical around 0, but
- Vary with $\mathbf{W}_i$

- Models for "balanced" scales (Jones & Sobel)
- Compound Ordered Hierarchical Probit ("`chopit`") (Wand & King)
- "Zero-Inflated" Ordered Models (Hill, Bagozzi, Moore & Mukherjee)
- Latent class/mixture models (Winkelmann, etc.)

Event Count Data

Event Count Properties

- Discrete / integer-values
- Non-negative
- “Cumulative”

$$\text{Arrival Rate} = \lambda$$

$$\Pr(\text{Event})_{t,t+h} = \lambda h$$

$$\Pr(\text{No Event})_{t,t+h} = 1 - \lambda h$$

$$\begin{aligned} \Pr(Y_t = y) &= \frac{\exp(-\lambda h)\lambda h^y}{y!} \\ &= \frac{\exp(-\lambda)\lambda^y}{y!} \end{aligned}$$

Poisson Assumptions

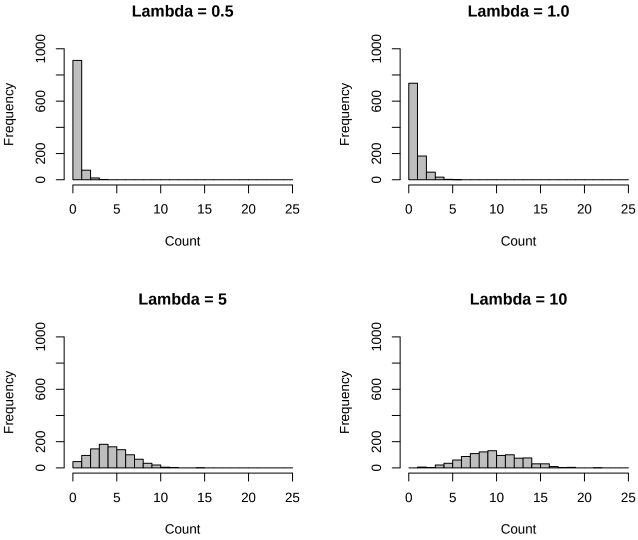
- No Simultaneous Events
- Constant Arrival Rate
- Independent Event Arrivals

For M independent Bernoulli trials with (sufficiently small) probability of success π and where $M\pi \equiv \lambda > 0$,

$$\begin{aligned} \Pr(Y_i = y) &= \lim_{M\to\infty} \left[\binom{M}{y} \left(\frac{\lambda}{M}\right)^{y} \left(1 - \frac{\lambda}{M}\right)^{M-y} \right] \\ &= \frac{\lambda^{y} \exp(-\lambda)}{y!} \end{aligned}$$

- Discrete
- $\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$
- Is not preserved under affine transformations...
- For $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$, $Z = X + Y \sim \text{Poisson}(\lambda_{X+Y})$ *iff* X and Y are *independent* but
- ...same is not true for differences.
- $\lambda \to \infty \iff Y \sim N$

Poissons: Examples



Suppose

$$E(Y_i) \equiv \lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$$

then

$$\Pr(Y_i = y|\mathbf{X}_i, \boldsymbol{\beta}) = \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^y}{y!}$$

$$L = \prod_{i=1}^{N} \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i}}{Y_i!}$$

$$\ln L = \sum_{i=1}^{N} [-\exp(\mathbf{X}_i\boldsymbol{\beta}) + Y_i\mathbf{X}_i\boldsymbol{\beta} - \ln(Y_i!)]$$

Poisson (and Negative Binomial) Regression

Poisson and negative binomial regression models for event counts were discussed in PLSC 503 (slides here and here; code here and here); that discussion included:

- Estimation / Model Fitting
- Interpretation:
  - Marginal Effects
  - Incidence Rate Ratios
  - Predicted Probabilities (including c.i.s and plots)
  - Predicted Counts
- Contagion, Heterogeneity, and Overdispersion
  - How event contagion can lead to over- (or sometimes under-) dispersion
  - Models for overdispersed (negative binomial) and underdispersed (continuous parameter binomial) event count data
  - Model fitting, interpretation, etc.

- Truncated Count Models
- Censored Count Models
- "Zero-Inflated" / "Hurdle" Count Models

Running Example: International Conflict(s)

- `conflicts` $= N$ of violent conflicts/year
- `polity` $=$ Rescaled POLITY IV democracy score
- `logPopulation` $= \ln(\text{population})$
- `logGDP` $= \ln(\text{GDP per capita})$
- `GDPGrowth` $=$ growth in GDP
- `logOpenness` $= \ln\left(\frac{\text{Imports + Exports}}{\text{GDP}}\right)$
- `govshareGDP` $=$ government's % of GDP

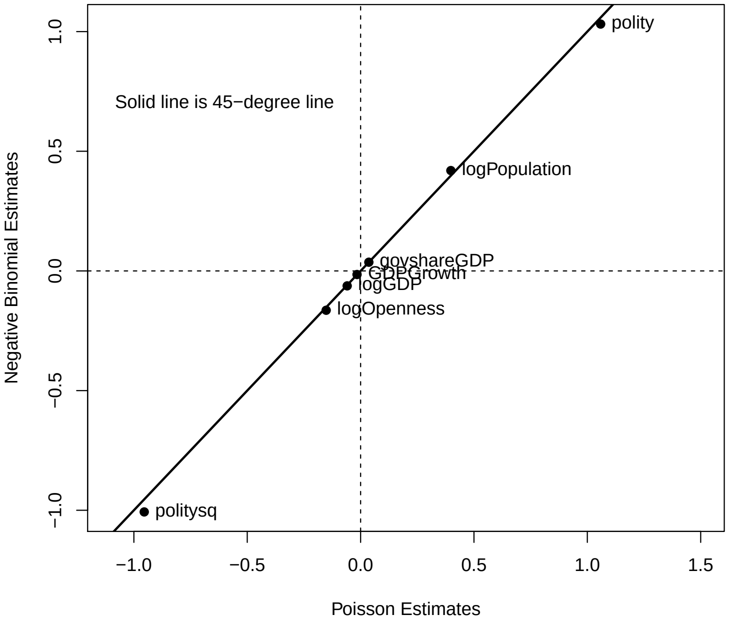
```
> summary(wars)
     ccode           year         conflicts     conflicts_no_zeros     polity     
 Min.   :  2   Min.   :1951   Min.   :0.000   Min.   :1           Min.   :0.000  
 1st Qu.:211   1st Qu.:1970   1st Qu.:0.000   1st Qu.:1           1st Qu.:0.150  
 Median :439   Median :1981   Median :0.000   Median :1           Median :0.450  
 Mean   :439   Mean   :1980   Mean   :0.304   Mean   :1           Mean   :0.527  
 3rd Qu.:640   3rd Qu.:1991   3rd Qu.:0.000   3rd Qu.:1           3rd Qu.:0.950  
 Max.   :950   Max.   :2000   Max.   :8.000   Max.   :8           Max.   :1.000  
                                              NA's   :4075                        
    politysq        population           GDP           openness       govshareGDP   
 Min.   :0.0000   Min.   :    122   Min.   :  171   Min.   :  3.7   Min.   : 2.97  
 1st Qu.:0.0225   1st Qu.:   3054   1st Qu.: 1401   1st Qu.: 30.9   1st Qu.:14.64  
 Median :0.2025   Median :   7725   Median : 3777   Median : 50.0   Median :18.94  
 Mean   :0.4278   Mean   :  33615   Mean   : 6641   Mean   : 62.2   Mean   :20.95  
 3rd Qu.:0.9025   3rd Qu.:  21979   3rd Qu.: 9032   3rd Qu.: 81.1   3rd Qu.:24.85  
 Max.   :1.0000   Max.   :1262474   Max.   :84408   Max.   :986.5   Max.   :83.68  

   GDPGrowth       logPopulation        logGDP        logOpenness   conflicts_censored
 Min.   :-63.32   Min.   : 4.80   Min.   : 5.14   Min.   :1.31   Min.   :0.000     
 1st Qu.: -0.90   1st Qu.: 8.02   1st Qu.: 7.25   1st Qu.:3.43   1st Qu.:0.000     
 Median :  2.08   Median : 8.95   Median : 8.24   Median :3.91   Median :0.000     
 Mean   :  1.92   Mean   : 8.99   Mean   : 8.23   Mean   :3.87   Mean   :0.299     
 3rd Qu.:  4.84   3rd Qu.:10.00   3rd Qu.: 9.11   3rd Qu.:4.40   3rd Qu.:0.000     
 Max.   :125.96   Max.   :14.05   Max.   :11.34   Max.   :6.89   Max.   :4.000     

    censored    
 Min.   :-1.00  
 1st Qu.: 1.00  
 Median : 1.00  
 Mean   : 0.99  
 3rd Qu.: 1.00  
 Max.   : 1.00  
```

```
> wars.poisson<-glm(conflicts~polity+politysq+logPopulation+logGDP+
  GDPGrowth+logOpenness+govshareGDP,family="poisson",data=wars)
> summary.glm(wars.poisson)

Call:
glm(formula = conflicts ~ polity + politysq + logPopulation +
    logGDP + GDPGrowth + logOpenness + govshareGDP, family = "poisson",
    data = wars)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.88565    0.36284  -13.47  < 2e-16 ***
polity         1.05866    0.39129    2.71   0.0068 **
politysq      -0.95432    0.37292   -2.56   0.0105 *
logPopulation  0.39809    0.01626   24.48  < 2e-16 ***
logGDP        -0.05919    0.02919   -2.03   0.0426 *
GDPGrowth     -0.01579    0.00345   -4.58  4.6e-06 ***
logOpenness   -0.15187    0.03691   -4.11  3.9e-05 ***
govshareGDP    0.03632    0.00235   15.48  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Basic Model: Negative Binomial

```
> wars.nb<-glm.nb(conflicts~polity+politysq+logPopulation+logGDP+GDPGrowth+
  logOpenness+govshareGDP,data=wars)
> summary(wars.nb)

Call:
glm.nb(formula = conflicts ~ polity + politysq + logPopulation +
    logGDP + GDPGrowth + logOpenness + govshareGDP, data = wars,
    init.theta = 2.10281397427423, link = log)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.987258   0.403221 -12.369  < 2e-16 ***
polity         1.031445   0.429147   2.403 0.016240 *
politysq      -1.006861   0.409911  -2.456 0.014038 *
logPopulation  0.419436   0.019065  22.000  < 2e-16 ***
logGDP        -0.062318   0.032646  -1.909 0.056276 .
GDPGrowth     -0.014965   0.003964  -3.775 0.000160 ***
logOpenness   -0.164250   0.041114  -3.995 6.47e-05 ***
govshareGDP    0.036494   0.002672  13.657  < 2e-16 ***
---
              Theta:  2.103
          Std. Err.:  0.322
```

Poisson ≈ Negative Binomial



Zero Truncation

$$\begin{aligned} \Pr(Y_i = 0) &= \frac{\exp(-\lambda_i)\lambda_i^0}{0!} \\ &= \exp(-\lambda_i) \end{aligned}$$

$$\Pr(Y_i > 0) = 1 - \exp(-\lambda_i).$$

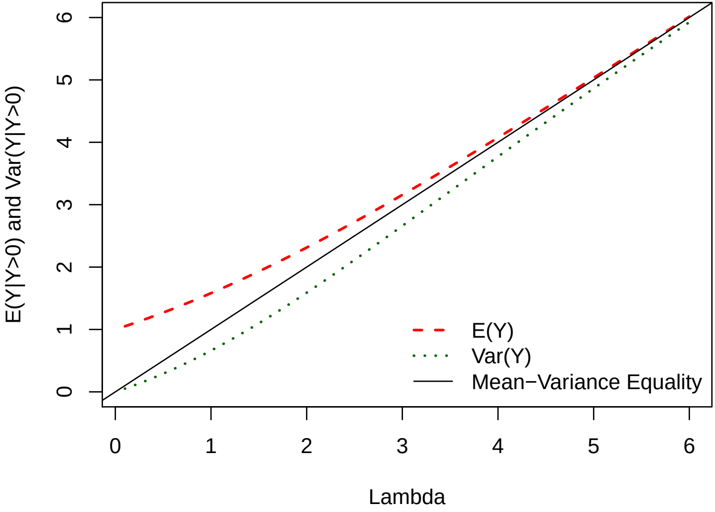
$$\begin{aligned} \Pr(Y_i = y | Y_i > 0) &= \frac{\Pr(Y_i = y)}{\Pr(Y_i > 0)} \\ &= \frac{\exp(-\lambda_i)\lambda_i^y}{y![1 - \exp(-\lambda_i)]} \end{aligned}$$

Zero Truncation (continued)

$$\mathsf{E}(Y|Y>0) = \frac{\lambda}{1-\exp(-\lambda)}$$

$$\begin{aligned}
\mathsf{Var}(Y|Y>0) &= \mathsf{E}(Y|Y>0) \times \{[1-\Pr(Y=0)]\,\mathsf{E}(Y|Y>0)\} \\
&= \frac{\lambda}{1-\exp(-\lambda)}\left[1-\frac{\lambda}{\exp(\lambda)-1}\right]
\end{aligned}$$

Zero Truncation Illustrated



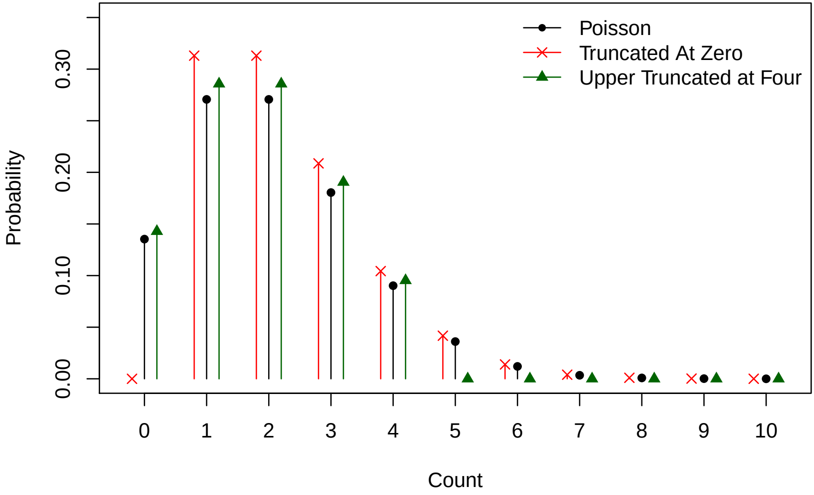
Upper Truncation

$$Y_i = \begin{cases} Y_i^* \text{ if } Y_i^* \leq \tau \\ \text{unobserved if } Y_i^* > \tau \end{cases}$$

$$\Pr(Y_i^* \leq \tau) = \sum_{y=0}^{\tau} \frac{\exp(-\lambda_i)\lambda_i^y}{y!}$$

$$\Pr(Y_i = y | Y_i \leq \tau) = \frac{\exp(-\lambda_i)\lambda_i^y}{y! \sum_{y=0}^{\tau} \frac{\exp(-\lambda_i)\lambda_i^y}{y!}}$$

Truncation Illustrated



Truncated Models: Estimation and Interpretation

$$\lambda_i = \exp(\mathbf{X}_i \boldsymbol{\beta})$$

- IRRs, predicted probabilities, etc. as usual
- Using formulae above

Zero-Truncated Models: (Incorrect/Poisson) Example

```
> wars.poisNo0s<-glm(conflicts_no_zeros~polity+politysq+logPopulation+
  logGDP+GDPGrowth+logOpenness+govshareGDP,family="poisson",data=wars)
> summary(wars.poisNo0s)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.4940523  0.3799176  -3.933 8.40e-05 ***
polity       -0.3508331  0.3858776  -0.909    0.363
politysq      0.4273003  0.3734399   1.144    0.253
logPopulation 0.1317254  0.0181912   7.241 4.45e-13 ***
logGDP        0.0389308  0.0320827   1.213    0.225
GDPGrowth    -0.0005765  0.0031788  -0.181    0.856
logOpenness  -0.0387960  0.0401573  -0.966    0.334
govshareGDP   0.0135720  0.0023168   5.858 4.68e-09 ***
---
    Null deviance: 396.43  on 1179  degrees of freedom
Residual deviance: 300.56  on 1172  degrees of freedom
  (11870 observations deleted due to missingness)
AIC: 2891.2

Number of Fisher Scoring iterations: 4
```

Zero-Truncated Models: Example

```
> library(VGAM)
> wars.0tpois<-vglm(conflicts_no_zeros~polity+politysq+logPopulation+
  logGDP+GDPGrowth+logOpenness+govshareGDP,pospoisson,data=wars)
> summary(wars.0tpois)

Coefficients:
                  Value Std. Error  t value
(Intercept)   -6.9985662  0.7802697 -8.96942
polity        -1.3061668  0.8185705 -1.59567
politysq       1.3202509  0.7876759  1.67613
logPopulation  0.3997250  0.0331791 12.04749
logGDP         0.2326896  0.0608548  3.82369
GDPGrowth     -0.0018478  0.0064828 -0.28503
logOpenness   -0.1045685  0.0779237 -1.34193
govshareGDP    0.0409683  0.0038646 10.60102

Number of linear predictors:  1
Name of linear predictor: log(lambda)
Dispersion Parameter for pospoisson family:   1
Log-likelihood: -806.6696 on 1172 degrees of freedom
Number of Iterations: 5
```

$$Y_i = \begin{cases} Y_i^* \text{ if } Y_i^* < k \\ k \text{ if } Y_i^* \geq k \end{cases}$$

$$\Pr(Y = y | Y^* < k) = \frac{\exp(-\lambda_i)\lambda_i^y}{y!},$$

$$\Pr(Y = k) = 1 - \sum_{y=0}^{k-1} \frac{\exp(-\lambda_i)\lambda_i^y}{y!}$$

Right Censoring

$$c_i = \begin{cases} 1 \text{ if } Y_i = k \\ 0 \text{ if } Y_i < k \end{cases}$$

$$\ln L = \sum_{i=1}^{N} (1 - c_i) \ln \left[\frac{\exp(-\lambda_i)\lambda_i^y}{y!} \right] + c_i \ln \left[1 - \sum_{y=0}^{k-1} \frac{\exp(-\lambda_i)\lambda_i^y}{y!} \right]$$

Left Censoring

$$Y_i = \begin{cases} \ell \text{ if } Y_i^* \leq \ell \\ Y_i^* \text{ if } Y_i^* > \ell \end{cases}$$

Double Censoring

$$Y_i = \begin{cases} \ell \text{ if } Y_i^* \leq \ell \\ Y_i^* \text{ if } \ell < Y_i^* < k \\ k \text{ if } Y_i^* \geq k \end{cases}$$

- R :
  - `vglm, pospoisson` (in `VGAM`) (zero truncation)
  - `vglm, cens.poisson` (in `VGAM`) (censored Poisson)

- Stata :
  - `ztp` / `ztnb` (zero truncation)
  - `trpoisson` (general truncation)
  - `cenpois` (censored Poisson)

Example, Again

$$c_i = \begin{cases} 1 \text{ if the observation's count is } \textit{uncensored}, \\ 0 \text{ if the observation's count is } \textit{left-censored}, \text{ and} \\ -1 \text{ if the observation's count is } \textit{right-censored}. \end{cases}$$

```
wars$censoredconflicts<-wars$conflicts
wars$censoredconflicts<-ifelse(wars$conflicts>3,4,wars$censoredconflicts)
wars$censindicator<-ifelse(wars$censoredconflicts==4,1,0)
```

Censored Example: (Incorrect) Poisson

```
> wars.poisCensored<-glm(censoredconflicts~polity+politysq+logPopulation+
  logGDP+GDPGrowth+logOpenness+govshareGDP,family="poisson",data=wars)
> summary(wars.poisCensored)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.743385   0.364843 -13.001  < 2e-16 ***
polity         1.070801   0.392156   2.731  0.00632 **
politysq      -1.025014   0.374139  -2.740  0.00615 **
logPopulation  0.381121   0.016491  23.111  < 2e-16 ***
logGDP        -0.049747   0.029444  -1.690  0.09111 .
GDPGrowth     -0.015869   0.003469  -4.574 4.78e-06 ***
logOpenness   -0.150489   0.037176  -4.048 5.16e-05 ***
govshareGDP    0.034396   0.002373  14.495  < 2e-16 ***
---
    Null deviance: 5007.5  on 5254  degrees of freedom
Residual deviance: 4059.2  on 5247  degrees of freedom
  (7795 observations deleted due to missingness)
AIC: 6644.6
Number of Fisher Scoring iterations: 6
```

Censored Example: Poisson

```
> wars.censpois<-vglm(SurvS4(censoredconflicts,censindicator)~polity+politysq+logPopulation
                     +logGDP+GDPGrowth+logOpenness+govshareGDP, cens.poisson,data=wars)
> summary(wars.censpois)

Call:
vglm(formula = SurvS4(censoredconflicts, censindicator) ~ polity +
    politysq + logPopulation + logGDP + GDPGrowth + logOpenness +
    govshareGDP, family = cens.poisson, data = wars)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.90350    1.22197    1.56    0.119
polity         1.65054    1.75924    0.94    0.348
politysq      -1.96215    1.67550   -1.17    0.242
logPopulation -0.05184    0.05040   -1.03    0.304
logGDP         0.07472    0.09577    0.78    0.435
GDPGrowth      0.00248    0.00809    0.31    0.759
logOpenness    0.09201    0.14403    0.64    0.523
govshareGDP   -0.01267    0.00741   -1.71    0.088 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of linear predictors:  1
Name of linear predictor: loge(mu)
Dispersion Parameter for cens.poisson family:   1
Log-likelihood: -51.72 on 5247 degrees of freedom
Number of iterations: 12
```

“Zero-Modified” Count Models

- “Zero-Inflated” Models
- “Hurdle” Models

"Zero-Inflated" Count Models

$$Y_i = p_i \times Y_i^*$$

$$\begin{aligned}
\Pr(Y_i = 0) &= \Pr(p_i = 0) + [\Pr(p_i = 1) \times \Pr(Y_i^* = 0)] \\
&= (1 - p_i^*) + p_i^*[\exp(-\lambda_i)]
\end{aligned}$$

$$\begin{aligned}
\Pr(Y_i = y) &= \Pr(p_i = 1) \times \Pr(Y_i^* = y) \\
&= p_i^* \times \frac{\exp(-\lambda_i)\lambda_i^y}{y!}
\end{aligned}$$

More on "Zero-Inflated" Models

$$\mathrm{E}(Y_i^*) \equiv \lambda_i = \exp(\mathbf{X}_i \boldsymbol{\beta})$$

$$\Pr(p_i = 1) \equiv p_i^* = \frac{1}{1 + \exp(-\mathbf{Z}_i \gamma)} \text{ or } \Phi(\mathbf{Z}_i \gamma)$$

- $\lambda_0 = \Pr(\text{No War}) = \exp(-\lambda)$
- $\lambda_+ = \Pr(Y \in \{1, 2, 3, ...\})$

$$\lambda_{0i} = \exp(\mathbf{X}_{0i}\boldsymbol{\beta}_0)$$

$$\lambda_{+i} = \exp(\mathbf{X}_{+i}\boldsymbol{\beta}_+)$$

Define:

$$\delta_i = \begin{cases} 0 \text{ if } Y_i = 0 \\ 1 \text{ if } Y_i > 0 \end{cases}$$

$$\begin{aligned} \ln L \quad = \quad & -\sum_{i=1}^{N} \delta_i \exp(\mathbf{X}_{0i}\boldsymbol{\beta}_0) + \sum_{i=1}^{N}(1-\delta_i)\{\ln[1-\exp(-\exp(\mathbf{X}_{0i}\boldsymbol{\beta}_0))] + \\ & Y_i(\mathbf{X}_{+i}\boldsymbol{\beta}_+) - \ln[\exp(\exp(\mathbf{X}_{+i}\boldsymbol{\beta}_+)) - 1]\} \end{aligned}$$

"Hurdle" Models: Details

$$\Pr(Y_i = 0) = 1 - \exp[-\exp(\mathbf{X}_{0i}\beta_0)]$$

- λ_+ defines a *truncated* Poisson process
- Y may be overdispersed, Poisson, or underdispersed

ZIP/ZINB and Hurdle Models: R

Command	Package	Count Distribution(s)	Transition Link(s)
Zero-Inflated Models			
`zeroinfl`	`pscl`	Poisson, NB, geometric	probit, logit, cloglog, log, Cauchy
`vglm,zipoisson`	`VGAM`	Poisson	logit, probit, cloglog, Cauchy, others
`vglm,zinegbinomial`	`VGAM`	Negative Binomial	logit, probit, cloglog, Cauchy, others
`cozigam`	`COZIGAM`	various	Various (see documentation)
Hurdle Models			
`hurdle`	`pscl`	Poisson, NB, geometric	binomial, Poisson, NB, geometric
`vglm,zapoisson`	`VGAM`	Poisson	logit, probit, cloglog, Cauchy, others

ZIP/ZINB and Hurdle Models: Stata

Command	Count Distribution	Transition Link(s)
Zero-Inflated Models		
`zip`	Poisson	probit or logit
`zinb`	Negative Binomial	probit or logit
Hurdle Models		
`hplogit`	Poisson	logit
`hpclg`	Poisson	complementary log-log
`hnblogit`	Negative Binomial	logit
`hnbclg`	Negative Binomial	complementary log-log

ZIP Example

```
wars.ZIP<-zeroinfl(conflicts~polity+politysq+logPopulation+logGDP+GDPGrowth+logOpenness+govshareGDP,
                   data=wars,dist="poisson",link="logit")
summary(wars.ZIP)

Call:
zeroinfl(formula = conflicts ~ polity + politysq + logPopulation +
    logGDP + GDPGrowth + logOpenness + govshareGDP, data = wars,
    dist = "poisson", link = "logit")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.4061 -0.5113 -0.3391 -0.0859 31.5392

Count model coefficients (poisson with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.79438    0.50297   -7.54  4.6e-14 ***
polity        -0.34467    0.52364   -0.66  0.51040
politysq       0.85585    0.50196    1.71  0.08819 .
logPopulation  0.27385    0.02336   11.72  < 2e-16 ***
logGDP        -0.14271    0.03993   -3.57  0.00035 ***
GDPGrowth     -0.00931    0.00405   -2.30  0.02138 *
logOpenness    0.20226    0.04963    4.08  4.6e-05 ***
govshareGDP    0.03138    0.00288   10.88  < 2e-16 ***
.
.
.
```

ZIP Example (continued)

```
.
.
.
Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.47668    1.80620    0.26  0.79185
polity        -3.97446    1.56375   -2.54  0.01103 *
politysq       5.34458    1.49540    3.57  0.00035 ***
logPopulation -0.62737    0.09152   -6.86  7.1e-12 ***
logGDP        -0.20497    0.13193   -1.55  0.12026
GDPGrowth      0.01898    0.01117    1.70  0.08933 .
logOpenness    1.73322    0.21124    8.21  2.3e-16 ***
govshareGDP   -0.03454    0.00889   -3.88  0.00010 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 31
Log-likelihood: -3.26e+03 on 16 Df
```

Example: Prose

- `polity`'s effect on the probability of being in the "zeros-only" state is curvilinear: it first decreases (as a country goes from being strongly autocratic to transitional) then increases (as it becomes more democratic).
- Growth and openness increase the probability of being in the zeros-only state, while government spending decreases it. e.g.:
  - A one-unit increase in `logOpenness` increases $\Pr(p_i = 0)$ by $(\exp(1.733)) \times 100 = 566$ percent.
  - Similarly, a one-unit (in this case, one-percent) increase in `govshareGDP` decreases $\Pr(p_i = 0)$ by $[1 - (\exp(-0.0345)) \times 100] = 3.4$ percent.
- A one-unit increase in `logOpenness` increases the incidence of armed conflicts by $(\exp(0.202)) \times 100 = 122$ percent.
- A one-unit increase in `logGDP`, by contrast, decreases the incidence of armed conflicts by $(1 - \exp(-0.1427)) \times 100 = 13.3$ percent.

Example: ZINB

```
wars.ZINB<-zeroinfl(conflicts~polity+politysq+logPopulation+logGDP+GDPGrowth+logOpenness+govshareGDP,
                    data=wars,dist="negbin",link="logit")
summary(wars.ZINB)

Call:
zeroinfl(formula = conflicts ~ polity + politysq + logPopulation +
    logGDP + GDPGrowth + logOpenness + govshareGDP, data = wars,
    dist = "negbin", link = "logit")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.2273 -0.5056 -0.3405 -0.0845 34.0184

Count model coefficients (negbin with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.86793    0.51463   -7.52  5.7e-14 ***
polity        -0.14355    0.54078   -0.27  0.79067
politysq       0.59083    0.52442    1.13  0.25990
logPopulation  0.28574    0.02491   11.47  < 2e-16 ***
logGDP        -0.14299    0.04100   -3.49  0.00049 ***
GDPGrowth     -0.00957    0.00416   -2.30  0.02151 *
logOpenness    0.17351    0.05298    3.28  0.00106 **
govshareGDP    0.03204    0.00313   10.24  < 2e-16 ***
Log(theta)     1.89010    0.37315    5.07  4.1e-07 ***
.
.
.
```

Example: ZINB (continued)

```
.
.
.
Zero-inflation model coefficients (binomial with logit link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.38441    1.98351    0.19  0.84633
polity        -3.66869    1.73171   -2.12  0.03413 *
politysq       5.05129    1.64892    3.06  0.00219 **
logPopulation -0.65973    0.09735   -6.78  1.2e-11 ***
logGDP        -0.23621    0.14249   -1.66  0.09737 .
GDPGrowth      0.01975    0.01181    1.67  0.09457 .
logOpenness    1.82769    0.22375    8.17  3.1e-16 ***
govshareGDP   -0.03515    0.00987   -3.56  0.00037 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Theta = 6.62
Number of iterations in BFGS optimization: 31
Log-likelihood: -3.25e+03 on 17 Df.
```

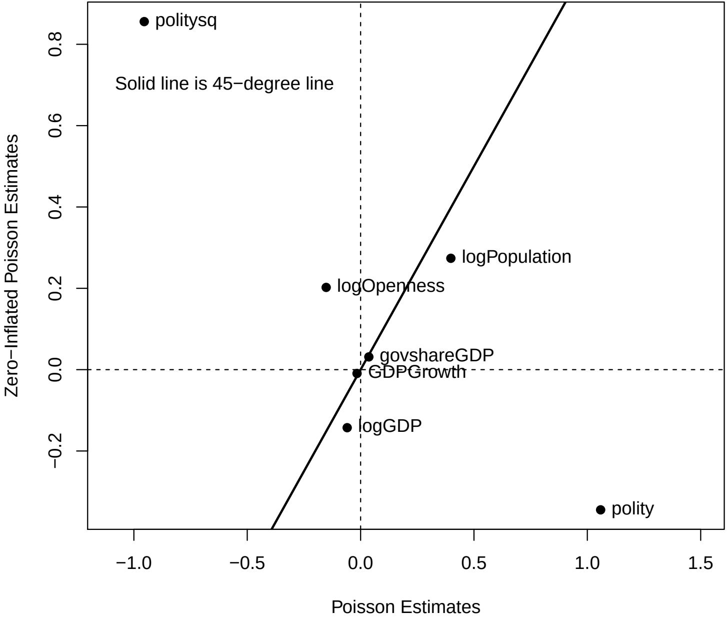
Example: Hurdle Poisson

```
> wars.hurdle<-hurdle(conflicts~polity+politysq+logPopulation+logGDP+GDPGrowth+
   logOpenness+govshareGDP,data=wars,dist=c("poisson"),zero.dist=c("poisson"),
   link=c("log"))
> summary(wars.hurdle)

Call:
hurdle(formula = conflicts ~ polity + politysq + logPopulation + logGDP + GDPGrowth +
    logOpenness + govshareGDP, data = wars, dist = c("poisson"), zero.dist = c("poisson"),
    link = c("log"))

Count model coefficients (truncated poisson with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.99856    0.78028   -8.97  < 2e-16 ***
polity        -1.30617    0.81861   -1.60  0.11058
politysq       1.32026    0.78772    1.68  0.09373 .
logPopulation  0.39973    0.03318   12.05  < 2e-16 ***
logGDP         0.23269    0.06085    3.82  0.00013 ***
GDPGrowth     -0.00185    0.00648   -0.29  0.77559
logOpenness   -0.10457    0.07793   -1.34  0.17963
govshareGDP    0.04097    0.00386   10.61  < 2e-16 ***

.
.
.
```

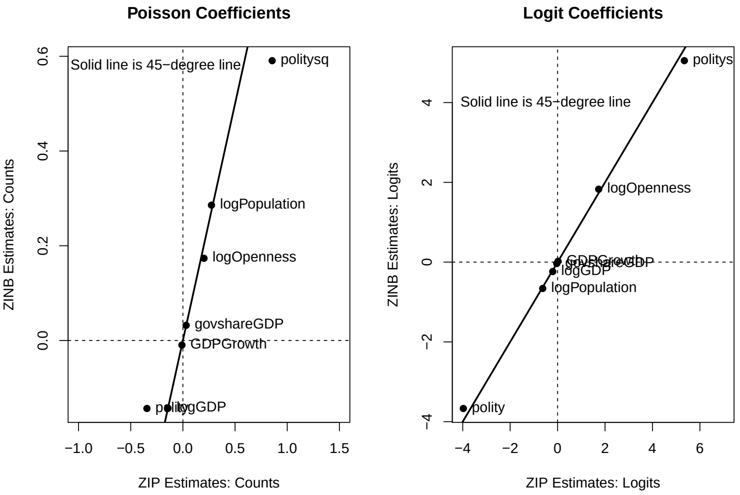
Example: Hurdle Poisson (continued)

```
.
.
.
Zero hurdle model coefficients (censored poisson with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.92477    0.41815   -9.39  < 2e-16 ***
polity         1.68586    0.44729    3.77  0.00016 ***
politysq      -1.61143    0.42744   -3.77  0.00016 ***
logPopulation  0.36963    0.01983   18.64  < 2e-16 ***
logGDP        -0.14103    0.03409   -4.14  3.5e-05 ***
GDPGrowth     -0.02152    0.00422   -5.09  3.5e-07 ***
logOpenness   -0.14945    0.04261   -3.51  0.00045 ***
govshareGDP    0.02909    0.00293    9.91  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 27
Log-likelihood: -3.27e+03 on 16 Df
```

Some More Comparisons: ZIP vs. Poisson



More Comparisons: ZIP vs. ZINB



Wrap-Up / Further Extensions

- Zero-Inflated Geometric Models (fixed degree of overdispersion; $\equiv$ negative binomial with $\alpha = 1$).

- There are Bayesian examples, too (see, e.g., the R package `bayescount`).

- Applications in survival analysis (more on that later...).

- Models for panel data with event count responses...