Harvard Department of Government 2003
Faraway Chapter 9, Other Generalized Linear Models

JEFF GILL
Visiting Professor, Fall 2024

Exponential Family Form Reminder

- The exponential family form of a PDF or PMF is:

$$f(z|\zeta) = \exp\big[t(z)u(\zeta)\big]r(z)s(\zeta)$$
$$= \exp\big[t(z)u(\zeta) + \log r(z) + \log s(\zeta)\big],$$

where: r and t are real-valued functions of z that do not depend on ζ, and s and u are real-valued functions of ζ that do not depend on z, and $r(z) > 0,\ s(\zeta) > 0\ \forall z, \zeta$.

Exponential Family Form Reminder

- The canonical form obtained by transforming: $y = t(z)$, and $\theta = u(\zeta)$. Call θ the canonical parameter. This produces the final form:

$$f(y|\theta) = \exp\big[y\theta - b(\theta) + c(y)\big].$$

- The exponential family form is invariant to sampling:

$$f(\mathbf{y}|\theta) = \exp\left[\sum y_i\theta - nb(\theta) + \sum c(y_i)\right].$$

- And there often exists a *scale parameter*:

$$f(\mathbf{y}|\theta) = \exp\left[\frac{\sum y_i\theta - nb(\theta)}{\phi} + \sum c(y_i, \phi)\right].$$

Exponential Family Form, Gamma Case

- Assume Y is distributed gamma indexed by two parameters: the shape parameter, and the inverse-scale parameter.

- The gamma distribution is most commonly written in "rate" format:

$$f(y|\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}y^{\alpha-1}e^{-\beta y}, \qquad y,\alpha,\beta > 0.$$

- `R` uses as a default the "scale" format:

$$f(y|\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^{-\alpha}y^{\alpha-1}e^{-y/\beta}, \qquad y,\alpha,\beta > 0.$$

- For our purposes a more convenient form is produced by transforming: $\alpha = \delta, \beta = \delta/\mu$, so

$$f(y|\mu,\delta) = \left(\frac{\delta}{\mu}\right)^{\delta}\frac{1}{\Gamma(\delta)}y^{\delta-1}\exp\left[\frac{-\delta y}{\mu}\right]$$

Exponential Family Form, Gamma Example

- Canonical form:

$$
\begin{aligned}
f(y|\mu,\delta) &= \left(\frac{\delta}{\mu}\right)^{\delta} \frac{1}{\Gamma(\delta)} y^{\delta-1} \exp\left[\frac{-\delta y}{\mu}\right] \\
&= \exp\left[\delta \log(\delta) - \delta \log(\mu) - \log(\Gamma(\delta)) + (\delta - 1)\log(y) - \frac{\delta y}{\mu}\right] \\
&= \exp\left[\Big(\underbrace{-\frac{1}{\mu}y}_{\theta y} - \underbrace{\log(\mu)}_{b(\theta)}\Big)\Big/\underbrace{\frac{1}{\delta}}_{a(\psi)} + \underbrace{\delta \log(\delta) + (\delta-1)\log(y) - \log(\Gamma(\delta))}_{c(y,\psi)}\right].
\end{aligned}
$$

- The canonical link for the gamma family variable μ, is $\theta = -\frac{1}{\mu}$.
- So $b(\theta) = \log(\mu) = \log\left(-\frac{1}{\theta}\right)$ with the restriction: $\theta < 0$. Therefore: $b(\theta) = -\log(-\theta)$.
- The χ^2 distribution is gamma$(\frac{\rho}{2}, \frac{1}{2})$ for ρ degrees of freedom, and the exponential distribution is gamma$(1, \beta)$.

Gamma GLM of Electoral Politics in Scotland

- On September 11, 1997 Scottish voters overwhelming (74.3%) approved the establishment of the first Scottish national parliament in nearly three hundred years.
- On the same ballot, the voters gave strong support (63.5%) to granting this parliament taxation powers.
- Data: 32 *Unitary Authorities* (also called council districts), U.K. government sources, includes 40 potential explanatory variables
- Used here: CouncilTax (COU), PerClaimantFemale (PCR), StdMortalityRatio (MOR), Active (ACT), GDP (GDP), Percentage5to15 (PER).

The model for these data using the gamma link function is produced by:

$$\begin{aligned}
\underbrace{g^{-1}(\boldsymbol{\theta})}_{32\times 1} &= g^{-1}(\boldsymbol{X\beta}) \\
&= -\frac{1}{\boldsymbol{X\beta}} \\
&= -\left[\mathbf{1}\beta_0 + \mathbf{COU}\beta_1 + \mathbf{PCR}\beta_2 + \mathbf{MOR}\beta_3 + \mathbf{ACT}\beta_4 + \mathbf{GDP}\beta_5\right]^{-1} \\
&= E[\mathbf{Y}] = E[\mathbf{YES}].
\end{aligned}$$

The systematic component here is $\boldsymbol{X\beta}$, the stochastic component is $\mathbf{Y} = \mathbf{YES}$, and the link function is $\boldsymbol{\theta} = -\frac{1}{\boldsymbol{\mu}}$.

Gamma GLM

```
scotland.df <-
  read.table("https://jeffgill.org/wp-content/uploads/2024/08/scotvote.dat_.txt",
  header=TRUE)
scottish.vote.glm <- glm((PerYesTax/100) ~ CouncilTax * PerClaimantFemale
                          + StdMortalityRatio + Active + GDP + Percentage5to15,
                          family=Gamma, data=scotland.df)
graph.summary(scottish.vote.glm)
Family: Gamma      Link function: inverse
                                Coef Std.Err. 0.95 Lower 0.95 Upper CIs:ZE+RO
(Intercept)                   -1.777    1.148     -4.026      0.473  |--o--|
CouncilTax                     0.005    0.002      0.002      0.008      |o|
PerClaimantFemale              0.203    0.053      0.099      0.308      |o|
StdMortalityRatio             -0.007    0.003     -0.012     -0.002      |o|
Active                         0.011    0.004      0.003      0.019      |o|
GDP                            0.000    0.000      0.000      0.000      |o|
Percentage5to15               -0.052    0.024     -0.099     -0.005      |o|
CouncilTax:PerClaimantFemale   0.000    0.000      0.000      0.000      |o|
N: 32    log-likelihood: 59.892    AIC: -111.784    Dispersion Parameter: 0.0035842
    Null deviance: 0.536 on 31 degrees of freedom
```

```
Residual deviance:  0.087 on 24 degrees of freedom
```

Exponential Family Form, Inverse Gaussian Case

- PDF:

$$f(x|\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right], \qquad x > 0, \mu > 0.$$

- Derive the exponential family form and identify $b(\theta)$ for the Inverse Gaussian distribution assumming $\lambda = 1$:

$$\begin{aligned} f(x|\mu,\lambda) &= \exp\left[\frac{1}{2}\log(1) - \frac{1}{2}\log(2\pi x^3) - \frac{1}{2\mu^2 x}(x-\mu)^2\right] \\ &= \exp\left[-\frac{1}{2}\log(2\pi) - \frac{3}{2}\log(x) - \frac{x}{2\mu^2} + \frac{1}{\mu} - \frac{1}{2x}\right] \\ &= \exp\left[\underbrace{-\frac{1}{2}\mu^{-2}x}_{\theta y} + \underbrace{\mu^{-1}}_{-b(\theta)} \underbrace{-\frac{1}{2}\log(2\pi) - \frac{3}{2}\log x - \frac{1}{2}x^{-1}}_{c(y)}\right] \end{aligned}$$

Exponential Family Form, Inverse Gaussian Case

- Since: $\theta = -\frac{1}{2}\mu^{-2}$, then $\mu^2 = (-2\theta)^{-1}$, $\mu = (-2\theta)^{-1/2}$.
- So: $b(\theta) = -1/\mu\big|_{\mu=1/\sqrt{-2\theta}} = -1/(-2\theta)^{-1/2} = -\sqrt{-2\theta}$, with $\theta < 0$.
- Derivatives:

$$\frac{d}{d\theta}(-(-2\theta)^{\frac{1}{2}}) = (-2\theta)^{-\frac{1}{2}}$$

and:

$$\frac{d^2}{d\theta^2}(-(-2\theta)^{\frac{1}{2}}) = -\frac{1}{2}(-2\theta)^{-\frac{3}{2}}$$

Inverse Gaussian GLM

```
scottish.vote2.glm <- glm((PerYesTax/100) ~ CouncilTax * PerClaimantFemale
                          + StdMortalityRatio + Active + GDP + Percentage5to15,
                         inverse.gaussian(link = "1/mu^2"), data=scotland.df)
graph.summary(scottish.vote2.glm)
Family: inverse.gaussian      Link function: 1/mu^2
                                  Coef Std.Err 0.95 Lower 0.95 Upper CIs:        ZE+RO
(Intercept)                    -10.726   3.884    -18.337     -3.114 |-----o-----|
CouncilTax                       0.019   0.005      0.008      0.030             |o|
PerClaimantFemale                0.771   0.182      0.415      1.127              |o|
StdMortalityRatio               -0.023   0.009     -0.040     -0.005             |o|
Active                           0.036   0.013      0.011      0.062             |o|
GDP                              0.000   0.000      0.000      0.000             |o|
Percentage5to15                 -0.172   0.079     -0.327     -0.018             |o|
CouncilTax:PerClaimantFemale -0.001   0.000     -0.001      0.000             |o|
N: 32    log-likelihood: 59.231    AIC: -110.461    Dispersion Parameter: 0.0061
    Null deviance: 0.891 on 31 degrees of freedom
Residual deviance:  0.15 on 24 degrees of freedom
```

Moments of the Exponential Family Form, Reminder

- Mean and Variance:

$$E[Y] = \mu = b'(\theta) \qquad\qquad var(Y) = b''(\theta)a(\phi)$$

- The mean is a function of θ only while the variance is a product of the location and the scale.
- The term $b''(\theta)$ is called the *variance function* and tells us how the the variance relates to the mean.
- For the normal,

$$b''(\theta) = \frac{\partial^2}{\partial\theta^2}b(\theta) = \frac{\partial^2}{\partial\theta^2}\theta^2/2 = \frac{\partial}{\partial\theta}\theta = 1$$

  meaning that the variance is independent of the mean (a special circumstance).
- Weighting of cases done with $a(\phi) = \phi/w_i$, where w_i is a known weight.

Joint Modeling of the Mean and Dispersion

- So far we have modeled mean effects, $\mu = E[Y]$ where the variance takes a known or assumed form $\mathrm{Var}[Y_i] = \phi V(\mu_i)$, and:
  - ϕ is the variance in a Gaussian model,
  - the squared coefficient of variation in the gamma model,
  - 1 in the binomial and Poisson models.
- Now we will let ϕ_i vary with some explanatory variables.
- Use the standard GLM definition of the mean (Faraway notation):

$$E[Y_i] = \mu_i \qquad \boldsymbol{\eta}_i = g(\boldsymbol{\mu}_i) = \sum_j x_{ij}\boldsymbol{\beta}_j \qquad \mathrm{Var}[Y_i] = \phi_i V(\mu_i) \qquad w_i = 1/\phi_i.$$

- Now model the dispersion parameter using d_i, an estimate of the dispersion:

$$E[d_i] = \phi_i \qquad \zeta_i = \log(\phi_i) = \sum_j z_{ij}\gamma_j \qquad \mathrm{Var}[d_i] = \tau\phi_i^2.$$

Joint Modeling of the Mean and Dispersion

- Note that the mean model produces the outcome, ϕ_i, for the dispersion model.
- And the dispersion model produces the weights, w_i, for the mean model.
- Typically the dispersion model uses a gamma specification, but this is not necessary.
- Often the variance defining covariates are a subset of the total covariates.
- Note that after this process the standard errors of the final model need to be bootstrapped.

Boring Welding Example, Run a Linear Model

```
library(faraway)
data(weldstrength)
lmod <- lm(Strength ~ Drying + Material + Preheating, weldstrength)
summary(lmod)

Residuals:
   Min     1Q Median     3Q    Max
-1.050 -0.200  0.075  0.106  1.100

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   43.625      0.262  166.25  < 2e-16
Drying         2.150      0.262    8.19  2.9e-06
Material      -3.100      0.262  -11.81  5.8e-08
Preheating    -0.375      0.262   -1.43     0.18

Residual standard error: 0.525 on 12 degrees of freedom
Multiple R-squared: 0.946,       Adjusted R-squared: 0.932
F-statistic: 69.6 on 3 and 12 DF,  p-value: 7.39e-08
```

Boring Welding Example, Build a Model of the Dispersion

- Use the squared studentized residuals:

$$\tilde{r}_i = \frac{(y_i - \hat{y}_i)^2}{1 - h_i}$$

as the response in the dispersion with a gamma GLM (log-link) and weights of $1 - h_i$ (actually it is h_{ii} as the diagonal of the hat matrix).

- Create the hat vector and the squared studentized residuals vector.

```
( h <- influence(lmod)$hat )
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25 0.25

( d <- residuals(lmod)^2/(1-h) )
        1         2         3         4         5         6         7         8
0.0075000 0.0033333 0.1008333 0.6533333 0.0133333 0.0208333 1.4700000 0.0075000
        9        10        11        12        13        14        15        16
0.0133333 0.1008333 0.1633333 0.0075000 0.1875000 0.0033333 0.0408333 1.6133333
```

Boring Welding Example, Gamma Dispersion Model

- Now run the dispersion model with weights `1-h` and outcome `d`.
- Then create the weights vector from the inverse of the fitted values.

```
gmod <- glm(d ~ Material+Preheating,family=Gamma(link=log),weldstrength,weights=1-h)

( w <- 1/fitted(gmod) )
      1       2       3       4       5       6       7       8       9      10
 2.9360 27.8687 45.1339  1.8129 27.8687  2.9360  1.8129 45.1339 27.8687  2.9360
     11      12      13      14      15      16
 1.8129 45.1339  2.9360 27.8687 45.1339  1.8129
```

Boring Welding Example, Gamma Dispersion Model Output

```
summary(gmod)
Deviance Residuals:
   Min       1Q   Median       3Q      Max
-2.063  -1.029  -0.766   0.260   1.747

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1.077      0.614   -1.75    0.103
Material      -2.733      0.709   -3.85    0.002
Preheating     0.482      0.709    0.68    0.509

(Dispersion parameter for Gamma family taken to be 1.5100)

    Null deviance: 43.193  on 15  degrees of freedom
Residual deviance: 21.286  on 13  degrees of freedom
AIC: -21.9

Number of Fisher Scoring iterations: 23
```

Boring Welding Example, Weighted Model To Account For Dispersion

```
lmod <- lm(Strength ~ Drying + Material + Preheating, weldstrength, weights=w)
summary(lmod)
Residuals:
   Min     1Q Median     3Q    Max
-1.651 -0.481  0.143  0.472  1.452

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  43.6698     0.2056  212.42  < 2e-16
Drying        1.9953     0.0969   20.58  1.0e-10
Material     -3.1355     0.2024  -15.49  2.7e-09
Preheating   -0.2433     0.0998   -2.44    0.031

Residual standard error: 0.855 on 12 degrees of freedom
Multiple R-squared: 0.982,        Adjusted R-squared: 0.978
F-statistic:  223 on 3 and 12 DF,  p-value: 8.67e-11
```

Note that `Preheating` is now statistically reliable.

Quasi-Likelihood

- Extends the GLM to cases where the parametric form of the likelihood is known to be misspecified (Wedderburn 1974).
- Instead of a full PDF/PMF, only the first two moments need to be specified.
- This creates a more flexible form that retains desireable GLM properties (i.e. those described in Fahrmeier and Kaufmann 1985 and Wedderburn 1976).

Defining Quasi-Likelihood

- Suppose that we know something about the parametric form of the distribution generating the data, but not in complete detail.
- Obviously this precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation.
- This estimation procedure only requires specification of the mean function of the data and a stipulated relationship between this mean function and the variance function.
- Suppose $\mathbf{Y}$ is the outcome vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\phi V(\boldsymbol{\mu})$.
- Since the components of $\mathbf{Y}$ are iid by assumption, $V(\boldsymbol{\mu})$ is a diagonal matrix where $V(\mu_i)$ only depends on μ_i.

Quasi-Likelihood Criteria

- Instead of taking the first derivative of log likelihood with respect to the parameter vector, $\boldsymbol{\theta}$, suppose we take this derivative with respect to the mean function in a generalized linear model, $\boldsymbol{\mu}$, with the analogous properties:
  - $E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = 0.$
  - $\text{Var}\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu_i}\right] = \frac{1}{\phi V(\mu_i)}.$
  - $-E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \mu_i^2}\right] = \frac{1}{\phi V(\mu_i)}$

  leaving the form of $\ell(\boldsymbol{\theta})$ vague for the moment.
- Therefore what we have here is a linkage between the mean function and the variance function that does not depend on the form of the likelihood function.
- This gives a replacement for the unknown specific form of the score function that still provides the neccessary properties for maximum likelihood estimation.
- We imitate these three criteria of the score function with a function that contains significantly less parametric information: only the mean and variance.

Quasi-Likelihood Details

- A "likelihood" function that satisfies these three conditions is:

$$q_i = \frac{\mathbf{y}_i - \mu_i}{\phi V(\mu_i)}$$

(McCullagh and Nelder 1989, p.325; Shao 1999, p.314).

- The associated contribution to the log likelihood function from the i^{th} point is defined by:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(\mu_i)} dt.$$

- This is comparable to the regular MLE construction:

$$x_1, x_1, \ldots, x_n \sim \text{ iid } f(x|\theta)$$

$$\ell(\theta|\mathbf{x}) = \log\left(f(x_1|\theta) f(x_2|\theta) \cdots f(x_n|\theta)\right) = \sum_{i=1}^{n} \log(f(x_i|\theta))$$

$$\frac{d}{d\theta}\ell(\theta|\mathbf{x}) \equiv 0 \longrightarrow \hat{\theta}$$

- With the substitution:

$$\log(f(x_i|\theta)) \Longrightarrow Q_i.$$

Quasi-Likelihood Details

- So finding the maximum likelihood estimator for this setup, $\hat{\boldsymbol{\theta}}$, is equivalent to solving:

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\sum_{i=1}^{n} Q_i &= \sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\left(\int_{y_i}^{\mu_i}\frac{y_i - t}{\phi V(\mu_i)}dt\right)\right] \\
&= \sum_{i=1}^{n}\frac{y_i - \mu_i}{\phi V(\mu_i)}\left(\frac{\partial\mu_i}{\partial\theta}\right) \\
&= \sum_{i=1}^{n}\frac{y_i - \mu_i}{\phi V(\mu_i)}\left(\frac{\mathbf{x}_i}{g(\mu_i)}\right) \\
&\equiv \mathbf{0},
\end{aligned}
$$

  where $g(\mu)$ is the canonical link function for a generalized linear model specification.

- We use the usual maximum likelihood engine for inference with complete asymptotic properties such as consistency and normality (McCullagh 1983), by only specifying the relationship between the mean and variance functions as well as the link function (which actually comes directly from the form of the outcome variable data).

Quasi-Likelihood, Example

- The easiest example assumes that the mean and variance function are related by:

$$\phi = \sigma^2 = 1, \text{ and } b(\theta(\mu_i)) = \frac{\theta(\mu_i)^2}{2},$$

  so that:

$$\mathrm{Var}(\mu) = \frac{\partial^2}{\partial\theta^2}\left[\frac{b(\theta(\mu_i))}{(\mu_i)^2}\right] = 1.$$

- Then it follows that:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi V(\mu)} dt = -\frac{(y_i - \mu_i)^2}{2}.$$

- The quasi-likelihood solution for $\hat{\boldsymbol{\theta}}$ comes from solving the quasi-likelihood equation:

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n} Q_i = \frac{\partial}{\partial\mu}\sum_{i=1}^{n}\left[-\frac{(y_i - \mu_i)^2}{2}\right] = -\sum_{i=1}^{n} y_i + n\mu \equiv \mathbf{0}.$$

- In other words, $\hat{\mu} = \bar{y}$, because this example was setup with the same assumptions as a normal maximum likelihood problem but without specifying a normal likelihood function.

Quasi-Likelihood, Extensions

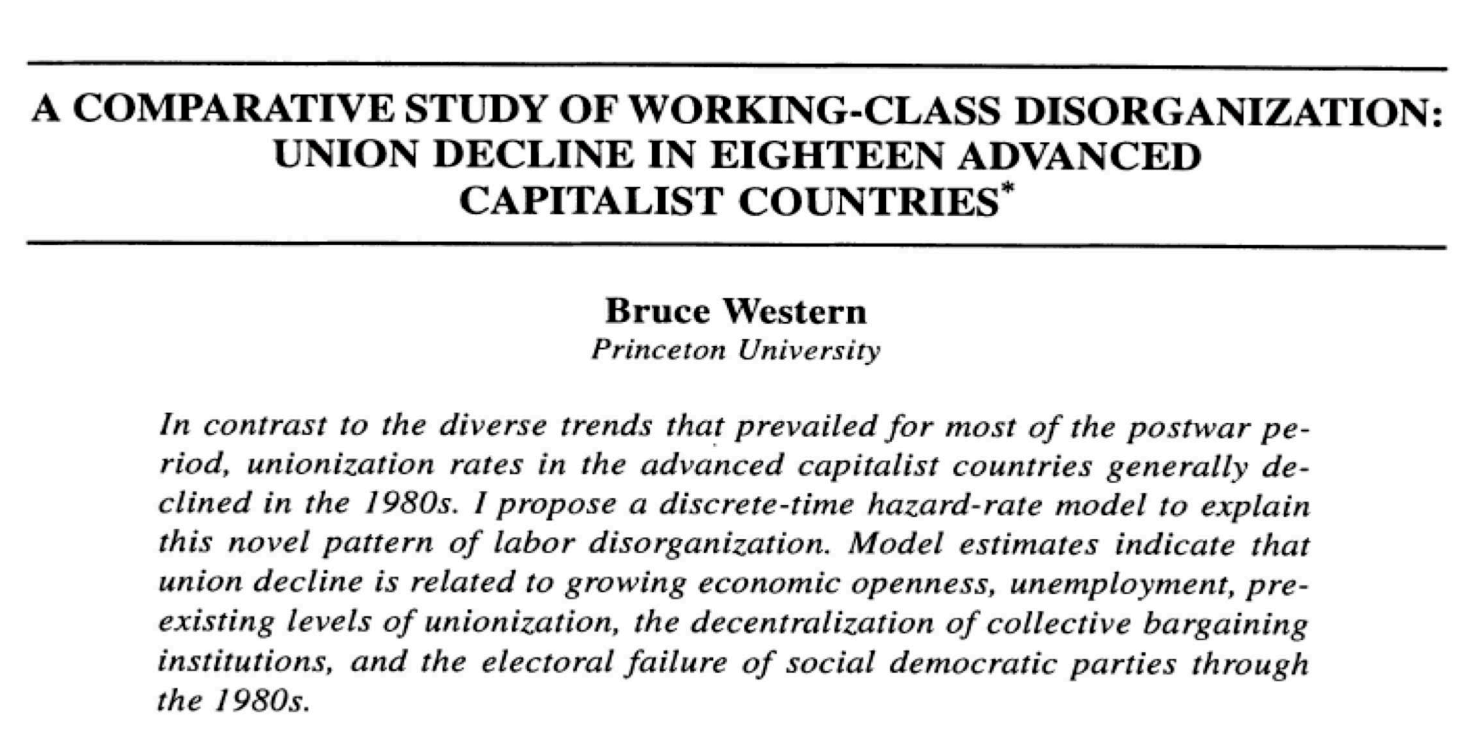
- Quasi-likelihood estimators are consistent, asymptotically equal to the true estimand (Fahrmeir and Tutz 2001, p.55-60, Firth 1987; McCullagh 1983).
- However, a quasi-likelihood estimator is often less efficient than a corresponding maximum likelihood estimator and can never be more efficient:

$$V_{\text{quasi}}(\theta) \geqslant [I(\theta)]^{-1},$$

where $I(\theta)$ is the Fisher information from the maximum likelihood estimation (McCullagh and Nelder 1987, p.347-8; Shao 1999, p.248-57).

- *Extended quasi-likelihood* models to compare different variance functions for the same data (Nelder and Pregibon 1987).
- *Pseudo-likelihood* models which build upon extended quasi-likelihood models by substituting a χ^2 component instead of a deviance component in dispersion analysis (Breslow 1990; Carroll and Ruppert 1982; Davidian and Carroll 1987).
- There are also models where the dispersion parameter is dependent on specified covariates (Smyth 1989).

Bruce Western, ASR 1995



A COMPARATIVE STUDY OF WORKING-CLASS DISORGANIZATION: UNION DECLINE IN EIGHTEEN ADVANCED CAPITALIST COUNTRIES*

Bruce Western
Princeton University

In contrast to the diverse trends that prevailed for most of the postwar period, unionization rates in the advanced capitalist countries generally declined in the 1980s. I propose a discrete-time hazard-rate model to explain this novel pattern of labor disorganization. Model estimates indicate that union decline is related to growing economic openness, unemployment, preexisting levels of unionization, the decentralization of collective bargaining institutions, and the electoral failure of social democratic parties through the 1980s.

Bruce Western, ASR 1995

Table 3. Year of Union Decline and Magnitude of Union Decline: 18 OECD Countries, 1973–1989

Country	Year of Union Decline	Magnitude of Decline[a]
High Union Density		
Belgium	1984	–.1
Denmark	1986	–.8
Finland	—	—
Sweden	—	—
Middle Union Density		
Australia	1982	–.3
Austria	1985	–.2
Canada	1982	–.1
Germany	1981	–.4
Ireland	1981	–.8
Italy	1980	–.4
New Zealand	1983	–.8
Norway	1983	–.3
United Kingdom	1980	–1.2
Low Union Density		
France	1977	–.4
Japan	1978	–.3
Netherlands	1979	–.6
Switzerland	1978	–.6
United States	1980	–.3

[a] Magnitude of decline is defined as the second difference of the smoothed union density time series in the year of union decline.

time can be written so all countries face a common baseline odds of union decline each year. Variability around this baseline is accounted for by the explanatory variables. Annual data are then collected from each country until the year of accelerating deunionization. Each country scores 0 on the dependent variable for each year, except the downturn year, which is coded 1. Countries with no union decline (Finland and Sweden) are censored, scoring 0 for all years. I obtain estimates of the effects of the explanatory variables from a logistic regression on the stacked time series collected for each country, using quasi-likelihood methods to estimate an extra component of dispersion in the dependent variable (McCullagh and Nelder 1989:124–28). I explore the quasi-likelihood fit by comparing its estimates to a robust fit that gives less weight to outlying observations (Pregibon 1982). This robust fit serves the dual purposes of providing a summary of the data that resists the influence of outliers while flagging outlying observations (with very small weights) for further analysis.

The dependence of the baseline hazard rate on time can be parameterized in several ways. The discussion of the economic sources of union decline suggests that pressures toward deunionization steadily increased after 1973. There is also evidence for the cross-national diffusion of employer tactics for opposing unionization (e.g., through American multinationals abroad). Increasing, but unmeasured, economic pressures and the diffusion of union opposition can be captured by an increasing baseline hazard rate. In this specification, the likelihood of a downturn in

Bruce Western, ASR 1995

Table 6. Quasi-Likelihood and Robust Coefficients from Discrete-Time Survival Analysis, Predicting Year of Union Decline: 18 OECD Countries, 1973–1989

Independent Variable	Quasi-Likelihood Coefficient (1)	Robust Coefficient (2)	Cross-Validation Bounds (3)	Extreme Bounds (4)
Constant	–2.42 (3.1)	–3.30 (3.2)	(–3.43, –1.93)	(–5.24, –.98)
Year	1.41 (6.4)	1.81 (5.6)	(1.32, 1.89)	(.24, 1.41)
Economic openness	.12 (3.1)	.17 (3.2)	(.08, .18)	(.04, .12)
Unemployment	1.18 (3.6)	1.54 (3.6)	(.96, 1.50)	(.89, 1.18)
Strike activity[a]	.11 (.04)	.08 (.28)	(–.55, .14)	(–1.50, 2.15)
Union density (lagged)	–.23 (6.0)	–.30 (5.4)	(–.30, –.21)	(–.23, –.11)
Decentralization	2.77 (4.3)	3.89 (4.7)	(2.18, 3.49)	(.33, 2.77)
Left Government	–4.41 (5.1)	–5.27 (4.7)	(–6.50, –4.04)	(–4.41, –1.71)
Dispersion	.35	.37		

Note: Numbers in parentheses under regression coefficients are absolute t-statistics.

[a]Coefficient has been multiplied by 10^4.

Chris Zorn, AJPS 2001

decision making in civil rights and liberties cases during three recent Court terms. I conclude with a discussion of the strengths and weaknesses of GEE models in general and an appendix on software issues.

Generalized Estimating Equation Models: An Overview

The GEE approach has its roots in the quasi-likelihood methods introduced by Wedderburn (1974) and Nelder and Wedderburn (1972) and developed and extended by McCullagh and Nelder (1983, 1989; see also Heyde 1997) and others.[3] While standard maximum-likelihood analysis specification of the full conditional distribution of the dependent variable, quasi-likelihood requires only that we postulate the relationship between the expected value of the outcome variable and the covariates and between the conditional mean and variance of the response vari-

[3]This section draws extensively on the presentation and notation of Zeger and Liang (1986) and Fitzmaurice, Laird, and Rotnitzky (1993). The literature on GEEs, particularly in biostatistics, is vast; good reviews of these models can be found in Liang, Zeger, and Qaqish (1992), Zeger and Liang (1992), Diggle, Liang, and Zeger (1994) and Ziegler, Kastner, and Blettner (1998).

ϕ

where ϕ is a scale parameter which may or may not be of substantive interest. The quasi-likelihood estimate of $\boldsymbol{\beta}$ is then the solution to a set of k "quasi-score" differential equations:

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathrm{D}_i' \mathrm{V}_i^{-1}\left(\mathrm{Y}_i - \mu_i\right) = 0 \qquad (3)$$

where $\mathbf{D}_i = \ \boldsymbol{\mu}_i / \ \boldsymbol{\beta}$. If the model is properly specified, then, asymptotically, $\mathrm{E}[\mathrm{U}_k(\boldsymbol{\beta})] = 0$ and $\mathrm{Cov}[\mathrm{U}_k(\boldsymbol{\beta})] = \mathbf{D}_i' V^{-1} \mathbf{D}_i$. The function $\mathrm{U}(\boldsymbol{\beta})$ thus behaves like the derivative of a log-likelihood (i.e., a score function); estimation may be accomplished either via generalized weighted least-squares or through an iterative process.[6]

[4]While this exposition is the standard one, it should be noted that the correlation within observations over time need not be temporal in nature; I address this further, and provide examples, below.

[5]For notational simplicity, I assume here that $T_i = T_{i'}, \forall\, i \;\; i'$, i.e., that the "panels" are balanced. This need not be the case for the models presented here; balanced panels are, however, necessary for likelihood-based "mixed parameter" models (Fitzmaurice, Laird, and Rotnitzky 1993).

[6]The standard reference for such models is McCullagh and Nelder (1989).

Jake Bowers, Political Psychology 2004

Calculating Effects

For an explanatory variable that takes on only values of 0 or 1 (such as Moved), we know that, holding all else constant,

$$\Pr(\widehat{y_{it}} = 1) = \begin{cases} \dfrac{e^{0\hat{\beta}}}{1+e^{0\hat{\beta}}} = \dfrac{1}{2}, \text{ when } x = 0 \\ \dfrac{e^{1\hat{\beta}}}{1+e^{1\hat{\beta}}} = \dfrac{e^{\hat{\beta}}}{1+e^{\hat{\beta}}}, \text{ when } x = 1 \end{cases} \tag{2}$$

The place on the logit function where a one-unit change in x is assumed to produce the most change in $\Pr(y)$ is at $\Pr(y) = .5$. Thus, calculating the effects with reference to a person who has a 50% probability of voting produces the maximum effect for a variable given the estimated coefficient. For example, in

[16] The particular estimation strategy I use here involved maximizing the restricted penalized quasi-likelihood function and was implemented in Splus via the glme command, provided to me as beta software by José C. Pinheiro. For technical details about the maximum likelihood estimation of generalized linear mixed-effects models like this one, see Breslow and Clayton (1993), Raudenbush and Bryk (2002, chapter 10), McCulloch and Searle (2001, chapter 8), and Snijders and Bosker (1999, chapter 14).

R Implementation of Quasilikelihood

- Add to our collection of link functions:

Binomial	`binomial(link=‘‘logit’’)`
Normal	`gaussian(link=‘‘identity’’)`
Gamma	`Gamma(link=‘‘inverse’’)`
Inverse Gamma	`inverse.gaussian(link =‘‘1/mu^2’’)`
Poisson	`poisson(link =‘‘log’’)`
Negative Binomial†	`negative.binomial(a=1,link=‘‘log’’)`
Quasi-Likelihood	`quasi(link=‘‘identity",variance=‘‘constant’’)`
Quasi-Likelihood/Binomial	`quasibinomial(link=‘‘logit’’)`
Quasi-Likelihood/Poisson	`quasipoisson(link=‘‘log’’)`

†Requires the Venables and Ripley `MASS` library extension.

- Quasi-Likelihood/Binomial and Quasi-Likelihood/poisson differ from Binomial and Poisson only in that the dispersion parameter is not fixed at one.

Australian Health Survey

- John Mullahy, "Heterogeneity, Excess Zeros, and the Structure of Count Data Models", Journal of Applied Econometrics, Vol. 12, No. 3, 1997, pp. 337-350.
- The data are in column-separated, multiple-lines-per-record, ASCII format, arrayed as follows:
    1. SEX
    2. AGE
    3. INCOME
    4. ILLNESS
    5. ACTDAYS
    6. HSCORE
    7. DOCTORCON
    8. LEVYPLUS
    9. FREEPOOR
    10. FREEREPAT
    11. CHCOND1
    12. CHCOND2
    13. AGESQ
- 5190 observations from 1977-1978 describing factors that affect health care utilization propensities.

Australian Health Survey

```
library(AER)
data("DoctorVisits")
dv_pois <- glm(visits ~ . + I(age^2), data = DoctorVisits, family = poisson)
summary(dv_pois)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-2.917  -0.686  -0.574  -0.484   5.701
```

Australian Health Survey

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.22385    0.18982  -11.72   <2e-16
genderfemale   0.15688    0.05614    2.79   0.0052
age            1.05630    1.00078    1.06   0.2912
income        -0.20532    0.08838   -2.32   0.0202
illness        0.18695    0.01828   10.23   <2e-16
reduced        0.12685    0.00503   25.20   <2e-16
health         0.03008    0.01010    2.98   0.0029
privateyes     0.12319    0.07164    1.72   0.0855
freepooryes   -0.44006    0.17981   -2.45   0.0144
freerepatyes   0.07980    0.09206    0.87   0.3860
nchronicyes    0.11409    0.06664    1.71   0.0869
lchronicyes    0.14116    0.08315    1.70   0.0896
I(age^2)      -0.84870    1.07778   -0.79   0.4310
(Dispersion parameter for poisson family taken to be 1)
    Null deviance: 5634.8  on 5189  degrees of freedom
Residual deviance: 4379.5  on 5177  degrees of freedom
AIC: 6737
```

Australian Health Survey

```
logLik(dv_pois)
'log Lik.' -3355.5 (df=13)

dv_nb <- glm.nb(visits ~ . + I(age^2), data = DoctorVisits)
summary(dv_nb)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.971  -0.635  -0.528  -0.441   4.007
```

Australian Health Survey

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.19001    0.23359   -9.38   <2e-16
genderfemale  0.21664    0.06970    3.11   0.0019
age          -0.21616    1.26670   -0.17   0.8645
income       -0.14220    0.10842   -1.31   0.1897
illness       0.21434    0.02358    9.09   <2e-16
reduced       0.14375    0.00731   19.66   <2e-16
health        0.03806    0.01365    2.79   0.0053
privateyes    0.11806    0.08581    1.38   0.1688
freepooryes  -0.49661    0.21080   -2.36   0.0185
freerepatyes  0.14498    0.11597    1.25   0.2112
nchronicyes   0.09935    0.07930    1.25   0.2103
lchronicyes   0.19033    0.10436    1.82   0.0682
I(age^2)      0.60916    1.38324    0.44   0.6597
(Dispersion parameter for Negative Binomial(0.9285) family taken to be 1)
    Null deviance: 3928.7  on 5189  degrees of freedom
Residual deviance: 3028.3  on 5177  degrees of freedom
AIC: 6425
```

Australian Health Survey

```
lrtest(dv_pois, dv_nb)
Likelihood ratio test

Model 1: visits ~ gender + age + income + illness + reduced + health +
    private + freepoor + freerepat + nchronic + lchronic + I(age^2)
Model 2: visits ~ gender + age + income + illness + reduced + health +
    private + freepoor + freerepat + nchronic + lchronic + I(age^2)
  #Df LogLik Df Chisq Pr(>Chisq)
1  13  -3356
2  14  -3199  1   314     <2e-16
```

Australian Health Survey

```
# TEST: VAR[y] = (1 + alpha) * mu = dispersion * mu.
dispersiontest(dv_pois)
        Overdispersion test

data:  dv_pois
z = 6.5428, p-value = 3.019e-11
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
    1.4144

dv_qpois <- glm(visits ~ . + I(age^2), data = DoctorVisits, family = quasipoisson)

summary(dv_qpois)
Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.917  -0.686  -0.574  -0.484   5.701
```

Australian Health Survey

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.2239	0.2187	-10.17	<2e-16
genderfemale	0.1569	0.0647	2.43	0.0153
age	1.0563	1.1532	0.92	0.3597
income	-0.2053	0.1018	-2.02	0.0438
illness	0.1870	0.0211	8.87	<2e-16
reduced	0.1268	0.0058	21.87	<2e-16
health	0.0301	0.0116	2.58	0.0098
privateyes	0.1232	0.0825	1.49	0.1357
freepooryes	-0.4401	0.2072	-2.12	0.0337
freerepatyes	0.0798	0.1061	0.75	0.4519
nchronicyes	0.1141	0.0768	1.49	0.1374
lchronicyes	0.1412	0.0958	1.47	0.1407
I(age^2)	-0.8487	1.2419	-0.68	0.4944

```
(Dispersion parameter for quasipoisson family taken to be 1.3278)
    Null deviance: 5634.8  on 5189  degrees of freedom
Residual deviance: 4379.5  on 5177  degrees of freedom
```

Other GLMs[40]