GLM Extensions

David Carlson

March 28, 2022

Extensions

- These are some of the more common extensions, but there are many more

Extensions

- These are some of the more common extensions, but there are many more
- Quasi-likelihoods

Extensions

- These are some of the more common extensions, but there are many more
- Quasi-likelihoods
- Generalized Linear Mixed-Effects Models (GLMM)

Extensions

- These are some of the more common extensions, but there are many more
- Quasi-likelihoods
- Generalized Linear Mixed-Effects Models (GLMM)
- Fractional regression

Extensions

- These are some of the more common extensions, but there are many more
- Quasi-likelihoods
- Generalized Linear Mixed-Effects Models (GLMM)
- Fractional regression
- Tobit models

Extensions

- These are some of the more common extensions, but there are many more
- Quasi-likelihoods
- Generalized Linear Mixed-Effects Models (GLMM)
- Fractional regression
- Tobit models
- Zero-inflated models

Quasi-Likelihood Overview

- Everyone pay attention to this slide

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified
- Precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation or a score function

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified
- Precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation or a score function
- Quasi-likelihood only requires specification of the mean function of the data and a stipulated relationship between this mean function and the variance function

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified
- Precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation or a score function
- Quasi-likelihood only requires specification of the mean function of the data and a stipulated relationship between this mean function and the variance function
- Table 7.1 has some common quasi-likelihoods

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified
- Precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation or a score function
- Quasi-likelihood only requires specification of the mean function of the data and a stipulated relationship between this mean function and the variance function
- Table 7.1 has some common quasi-likelihoods
- Quasi-likelihood estimator is often less efficient than MLE and can never be more efficient

Quasi-Likelihood Overview

- Everyone pay attention to this slide
- There are cases when there is not enough information about the distribution of the data, or the parametric form of the likelihood is known to be misspecified
- Precludes the standard maximum likelihood estimation of unknown parameters since we cannot specify a full likelihood equation or a score function
- Quasi-likelihood only requires specification of the mean function of the data and a stipulated relationship between this mean function and the variance function
- Table 7.1 has some common quasi-likelihoods
- Quasi-likelihood estimator is often less efficient than MLE and can never be more efficient
- E.g., quasi-Poisson: When there is overdispersion, allows us to model the variance as a linear function of the mean in contrast to the underlying assumption of a Poisson model that $\mu = \tau^2$ (can account for outliers)

Logit Intuition

- Recall the latent variable approach to the (binary) logit

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0, \\ 1 & \text{if} \quad 0 \leq y^* \end{cases}$$

- We do not know the latent variable by definition

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0, \\ 1 & \text{if} \quad 0 \leq y^* \end{cases}$$

- We do not know the latent variable by definition
- We cannot assess the fit (except for train / test split)

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0, \\ 1 & \text{if} \quad 0 \leq y^* \end{cases}$$

- We do not know the latent variable by definition
- We cannot assess the fit (except for train / test split)
- We cannot assume a distribution

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0, \\ 1 & \text{if} \quad 0 \leq y^* \end{cases}$$

- We do not know the latent variable by definition
- We cannot assess the fit (except for train / test split)
- We cannot assume a distribution
- It is often equally or very similarly fit with various assumptions about the linking distribution

Logit Intuition

- Recall the latent variable approach to the (binary) logit
- Imagine a latent variable $y^* = X\beta + \epsilon$
- We can reformulate our understanding into a decision rule:

$$y = \begin{cases} 0 & \text{if} \quad y^* < 0, \\ 1 & \text{if} \quad 0 \leq y^* \end{cases}$$

- We do not know the latent variable by definition
- We cannot assess the fit (except for train / test split)
- We cannot assume a distribution
- It is often equally or very similarly fit with various assumptions about the linking distribution
- At the extreme, imagine perfect separability along x; a large number of fits will perfectly predict the data

Quasi-Likelihood Estimation

- Only potential ("hardcore") methodologists need to pay attention

Quasi-Likelihood Estimation

- Only potential ("hardcore") methodologists need to pay attention
- Quasi-score function: $q_i = \frac{y_i - \mu_i}{a(\psi)\tau^2}$
- Contribution of ith point to log-likelihood function: $Q_i = \int_{y_i}^{\mu_i} \frac{y_i - \mu_i}{a(\psi)\tau^2} dt$
- Components of $\mathbf{Y}$ are independent by assumption (we can violate this in later weeks), the log-quasi-likelihood for the complete data is the sum of the individual contributions: $Q(\theta, a(\psi)|y) = \sum_{i=1}^{n} Q_i$
- MLE of $\hat{\theta}$: $\frac{\partial}{\partial\theta} Q(\theta, \psi|y) = -\sum_{i=1}^{n} y_i + n\theta \equiv 0$
- Quasi-deviance function:

$$D(\theta, \psi|y) = -2a(\psi)^{-1} \sum_{i=1}^{n} Q_i = 2\int_{\mu_i}^{y_i} \frac{y_i^{-t}}{\tau^2} dt$$

Generalized Linear Mixed-Effects Model (GLMM)

- So far, although rarely referred to in this way unless using TSCS, we have only been dealing with fixed-effects

Generalized Linear Mixed-Effects Model (GLMM)

- So far, although rarely referred to in this way unless using TSCS, we have only been dealing with fixed-effects
- We will deal with in more detail in the TSCS week and in Bayesian

Generalized Linear Mixed-Effects Model (GLMM)

- So far, although rarely referred to in this way unless using TSCS, we have only been dealing with fixed-effects
- We will deal with in more detail in the TSCS week and in Bayesian
- Mixed-effects models consider the dependencies of the observations within clusters and allow us not only to reach unbiased estimates of the effect of covariates of interest and their respective standard errors but also to address questions related to the variation between and within groups: Analyze the trajectories of groups/individuals through time, assess the differences between clusters, and others

Generalized Linear Mixed-Effects Model (GLMM)

- So far, although rarely referred to in this way unless using TSCS, we have only been dealing with fixed-effects
- We will deal with in more detail in the TSCS week and in Bayesian
- Mixed-effects models consider the dependencies of the observations within clusters and allow us not only to reach unbiased estimates of the effect of covariates of interest and their respective standard errors but also to address questions related to the variation between and within groups: Analyze the trajectories of groups/individuals through time, assess the differences between clusters, and others
- This approach is useful for panel data where responses recorded through time are perfectly grouped by panelist

Generalized Linear Mixed-Effects Model (GLMM)

- So far, although rarely referred to in this way unless using TSCS, we have only been dealing with fixed-effects
- We will deal with in more detail in the TSCS week and in Bayesian
- Mixed-effects models consider the dependencies of the observations within clusters and allow us not only to reach unbiased estimates of the effect of covariates of interest and their respective standard errors but also to address questions related to the variation between and within groups: Analyze the trajectories of groups/individuals through time, assess the differences between clusters, and others
- This approach is useful for panel data where responses recorded through time are perfectly grouped by panelist
- For GLMMs, we add random effects to the linear predictor and then express the expected value of the outcome conditional on those random effects

Random Effects Intro

- Effect of being a unit of observation (if the random effect is at the unit-level)

Random Effects Intro

- Effect of being a unit of observation (if the random effect is at the unit-level)
- If the subjects in our sample have been chosen randomly with the goal of treating them as a representation of the population of interest, then their effects on the outcome are also going to be random and generalizable to that same population

Random Effects Intro

- Effect of being a unit of observation (if the random effect is at the unit-level)
- If the subjects in our sample have been chosen randomly with the goal of treating them as a representation of the population of interest, then their effects on the outcome are also going to be random and generalizable to that same population
- Random variable that not only will help to make inferences about the population but also allows us to assess the variation between individuals, predict outcomes for each of them, and incorporate the existent correlation between observations

Random Effects Intro

- Effect of being a unit of observation (if the random effect is at the unit-level)
- If the subjects in our sample have been chosen randomly with the goal of treating them as a representation of the population of interest, then their effects on the outcome are also going to be random and generalizable to that same population
- Random variable that not only will help to make inferences about the population but also allows us to assess the variation between individuals, predict outcomes for each of them, and incorporate the existent correlation between observations
- There is therefore a distributional assumption on random effects (as opposed to with fixed effects)

Random Effects Intro

- Effect of being a unit of observation (if the random effect is at the unit-level)
- If the subjects in our sample have been chosen randomly with the goal of treating them as a representation of the population of interest, then their effects on the outcome are also going to be random and generalizable to that same population
- Random variable that not only will help to make inferences about the population but also allows us to assess the variation between individuals, predict outcomes for each of them, and incorporate the existent correlation between observations
- There is therefore a distributional assumption on random effects (as opposed to with fixed effects)
- Generally more power than fixed effects, but need to make the above assumptions, because with greater power generally comes larger false positive rates if the assumptions are not met

Fractional Regression

- Consider the logit example above

Fractional Regression

- Consider the logit example above
- For proportions, we could use quasi-binomial

Fractional Regression

- Consider the logit example above
- For proportions, we could use quasi-binomial
- But there are some undesirable properties of the estimator (such as the fact that proportions rarely follow the specified distribution)

Fractional Regression

- Consider the logit example above
- For proportions, we could use quasi-binomial
- But there are some undesirable properties of the estimator (such as the fact that proportions rarely follow the specified distribution)
- Examples of proportional data (not always obvious)?

The Tobit Model

- We often have censoring in the outcome

The Tobit Model

- We often have censoring in the outcome
- Censoring can be on either or both sides

The Tobit Model

- We often have censoring in the outcome
- Censoring can be on either or both sides
- Example measurements that are censored?

The Tobit Model

- We often have censoring in the outcome
- Censoring can be on either or both sides
- Example measurements that are censored?
- The Tobit allows for and accounts for DGPs that would have created larger (or smaller) observations, but the censorship prohibits the actual observation above (or below) a threshold

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model
- Original ZI: The logit

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model
- Original ZI: The logit
  - Turns out, if you have a bunch of zeros, you can simply drop a lot of them (but maintain all ones)

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model
- Original ZI: The logit
  - Turns out, if you have a bunch of zeros, you can simply drop a lot of them (but maintain all ones)
  - The only thing that will change is the intercept term, which we only care about for prediction

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model
- Original ZI: The logit
  - Turns out, if you have a bunch of zeros, you can simply drop a lot of them (but maintain all ones)
  - The only thing that will change is the intercept term, which we only care about for prediction
  - Slope estimates (what we usually care about) remain unchanged

Zero-Inflated Models

- When there is an excess of zeros, we may consider a ZI model
- Original ZI: The logit
  - Turns out, if you have a bunch of zeros, you can simply drop a lot of them (but maintain all ones)
  - The only thing that will change is the intercept term, which we only care about for prediction
  - Slope estimates (what we usually care about) remain unchanged
  - Try to prove this mathematically, or simply run some examples to assure yourself this is the case

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression
- Model the probability that the observation is an always zero vs. a potential non-zero

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression
- Model the probability that the observation is an always zero vs. a potential non-zero
- Then multiply the probability of a potential non-zero with the distribution of interest

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression
- Model the probability that the observation is an always zero vs. a potential non-zero
- Then multiply the probability of a potential non-zero with the distribution of interest
- Could be logit, Probit, Poisson, ZIMVOP, etc.

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression
- Model the probability that the observation is an always zero vs. a potential non-zero
- Then multiply the probability of a potential non-zero with the distribution of interest
- Could be logit, Probit, Poisson, ZIMVOP, etc.
- Hurdle models also allow for undercount of zeros

More Advanced (and interesting) ZI Models

- More recent developments involve two-stage (but simultaneously estimated) regression
- Model the probability that the observation is an always zero vs. a potential non-zero
- Then multiply the probability of a potential non-zero with the distribution of interest
- Could be logit, Probit, Poisson, ZIMVOP, etc.
- Hurdle models also allow for undercount of zeros
- Again, we will cover in more detail in Bayesian weeks