

Robust diagnostics for count regression models

by Tsung-Chi Cheng

Discussant: Philipp Pfeiffer

CRC 649 - Motzen 2014

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion
- ▶ However, often analyzed incorrectly when structurally estimating a distribution

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion
- ▶ However, often analyzed incorrectly when structurally estimating a distribution
- ▶ We need robust estimators that can correctly identify outliers

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion
- ▶ However, often analyzed incorrectly when structurally estimating a distribution
- ▶ We need robust estimators that can correctly identify outliers
- ▶ George proposes a new minimum trimmed distance estimator

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion
- ▶ However, often analyzed incorrectly when structurally estimating a distribution
- ▶ We need robust estimators that can correctly identify outliers
- ▶ George proposes a new minimum trimmed distance estimator
 - ▶ Focus on likelihood function instead of data trimming

Summary

- ▶ Count data common in many fields (e.g., # of world cup goals by player, # of hospital visits, # of banks going bust)
- ▶ Typical problems: excess number of zero and overdispersion
- ▶ However, often analyzed incorrectly when structurally estimating a distribution
- ▶ We need robust estimators that can correctly identify outliers
- ▶ George proposes a new minimum trimmed distance estimator
 - ▶ Focus on likelihood function instead of data trimming
- ▶ Superiority of this approach is illustrated by examples and simulation studies

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?
- ▶ What drives the results?
 - ▶ Closer look into the theory could be helpful

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?
- ▶ What drives the results?
 - ▶ Closer look into the theory could be helpful
- ▶ Slightly confusing use of R functions
 - ▶ To outsiders a little unclear what algorithms and estimators are used and how they are connected

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?
- ▶ What drives the results?
 - ▶ Closer look into the theory could be helpful
- ▶ Slightly confusing use of R functions
 - ▶ To outsiders a little unclear what algorithms and estimators are used and how they are connected
 - ▶ What is the relation between MTLE and MTDE?

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?
- ▶ What drives the results?
 - ▶ Closer look into the theory could be helpful
- ▶ Slightly confusing use of R functions
 - ▶ To outsiders a little unclear what algorithms and estimators are used and how they are connected
 - ▶ What is the relation between MTLE and MTDE?
- ▶ Clearly highlight the contribution

Discussion

- ▶ Striking results! The proposed methodology can be used for robustly fitting mixture models
- ▶ What role does the algorithm choice play?
 - ▶ Can subset sampling be improved?
 - ▶ Other algorithms applicable?
- ▶ What drives the results?
 - ▶ Closer look into the theory could be helpful
- ▶ Slightly confusing use of R functions
 - ▶ To outsiders a little unclear what algorithms and estimators are used and how they are connected
 - ▶ What is the relation between MTLE and MTDE?
- ▶ Clearly highlight the contribution
- ▶ Offer a little road map