Discussant for Llobet and Gauthier et al.

Max Goplerud

University of Texas at Austin

EPSA (2025)

Llobet: Overview

- "Bad" common practice by those who create our data:
    - Censoring continuous z into K discrete categories?

Llobet: Overview

- "Bad" common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., "midpoint") to complex (e.g., "rank likelihood")

Llobet: Overview

- "Bad" common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., "midpoint") to complex (e.g., "rank likelihood")
- At the moment, simulations only $\rightarrow$ an application

Llobet: Overview

- "Bad" common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., "midpoint") to complex (e.g., "rank likelihood")
- At the moment, simulations only $\rightarrow$ an application
  - Do this next! Can reveal interesting wrinkles
  - Find one that is **isn't** income or refocus the paper on income?

Llobet: Overview

- “Bad” common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., “midpoint”) to complex (e.g., “rank likelihood”)
- At the moment, simulations only $\rightarrow$ an application
  - Do this next! Can reveal interesting wrinkles
  - Find one that is **isn’t** income or refocus the paper on income?
- Focus on bias in coefficient of unobserved thing $\rightarrow$ tricky

Llobet: Overview

- "Bad" common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., "midpoint") to complex (e.g., "rank likelihood")
- At the moment, simulations only $\rightarrow$ an application
  - Do this next! Can reveal interesting wrinkles
  - Find one that is **isn't** income or refocus the paper on income?
- Focus on bias in coefficient of unobserved thing $\rightarrow$ tricky
  - If $R \in \{1, 2, 3, 4, 5\}$ and I multiplied by 100, so what?

Llobet: Overview

- “Bad” common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., “midpoint”) to complex (e.g., “rank likelihood”)
- At the moment, simulations only $\rightarrow$ an application
  - Do this next! Can reveal interesting wrinkles
  - Find one that is **isn’t** income or refocus the paper on income?
- Focus on bias in coefficient of unobserved thing $\rightarrow$ tricky
  - If $R \in \{1, 2, 3, 4, 5\}$ and I multiplied by 100, so what?
  - Focus on t-stat (or power, etc.) vs. bias

Llobet: Overview

- "Bad" common practice by those who create our data:
  - Censoring continuous z into K discrete categories?
- Careful and readable discussion of the different strategies one might do–from simple (e.g., "midpoint") to complex (e.g., "rank likelihood")
- At the moment, simulations only $\rightarrow$ an application
  - Do this next! Can reveal interesting wrinkles
  - Find one that is **isn't** income or refocus the paper on income?
- Focus on bias in coefficient of unobserved thing $\rightarrow$ tricky
  - If $R \in \{1, 2, 3, 4, 5\}$ and I multiplied by 100, so what?
  - Focus on t-stat (or power, etc.) vs. bias
  - Leads to a puzzle about the 'continuous" strategy

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
  - Orders observations from $1, \cdots, N$ (not unreasonable...)

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
  - Orders observations from $1, \cdots, N$ (not unreasonable...)
  - But, this gives a very small coefficient as the **scale** is massive! (Variance $\approx N^2/12$)

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
  - Orders observations from $1, \cdots, N$ (not unreasonable...)
  - But, this gives a very small coefficient as the **scale** is massive! (Variance $\approx N^2/12$)
  - Rescale it by the $sd(Z)$ (or sd. of mid-point), bias goes away

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
  - Orders observations from $1, \cdots, N$ (not unreasonable...)
  - But, this gives a very small coefficient as the **scale** is massive! (Variance $\approx N^2/12$)
  - Rescale it by the $sd(Z)$ (or sd. of mid-point), bias goes away
    - If rescale by mid-point $\rightarrow$ basically identical

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
  - Orders observations from $1, \cdots, N$ (not unreasonable...)
  - But, this gives a very small coefficient as the **scale** is massive! (Variance $\approx N^2/12$)
  - Rescale it by the $sd(Z)$ (or sd. of mid-point), bias goes away
    - If rescale by mid-point $\rightarrow$ basically identical
  - Rescaling is arbitrary so inference (t-stat, etc.) is same with/out

Llobet: Avoid Treating R As Continuous

- "The findings here provide a clear warning against the widespread, but misguided, practice, often cited in the psychometrics literature, of treating ordinal variables as continuous when $k > 5$."
- A (partial) defense of the psychometricians:
- Consider the "limiting case" for your scenario: If we have N observations, imagine we had $K = N$ categories.
    - Orders observations from $1, \cdots, N$ (not unreasonable...)
    - But, this gives a very small coefficient as the **scale** is massive! (Variance $\approx N^2/12$)
    - Rescale it by the $sd(Z)$ (or sd. of mid-point), bias goes away
        - If rescale by mid-point $\rightarrow$ basically identical
    - Rescaling is arbitrary so inference (t-stat, etc.) is same with/out
- If the coefficient's magnitude is of interest $\rightarrow$ rescale!

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
    - How is uncertainty in $\hat{Z}$ transferred into regression stage?

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)
  - No direct effects of X on Y?
  - Many real world settings would have this (see Knox et al.)

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)
  - No direct effects of X on Y?
  - Many real world settings would have this (see Knox et al.)
- Choice of Distribution: What happens if we get it wrong?

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)
  - No direct effects of X on Y?
  - Many real world settings would have this (see Knox et al.)
- Choice of Distribution: What happens if we get it wrong?
  - Z is log-normal but we say Pareto, Z is Pareto but we say log-normal

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)
  - No direct effects of X on Y?
  - Many real world settings would have this (see Knox et al.)
- Choice of Distribution: What happens if we get it wrong?
  - Z is log-normal but we say Pareto, Z is Pareto but we say log-normal
- "extremely heavy-tailed distributions like the Pareto, all methods exhibit limitations"

Llobet: Other Thoughts

- Method of Composition: More clarity on error propagation
  - How is uncertainty in $\hat{Z}$ transferred into regression stage?
  - Trier and Jackman (2008) or Knox, Lucas, and Tam Cho (2022)
- Simulations: Regression (γ) separately from that on predicting Z (β)
  - No direct effects of X on Y?
  - Many real world settings would have this (see Knox et al.)
- Choice of Distribution: What happens if we get it wrong?
  - Z is log-normal but we say Pareto, Z is Pareto but we say log-normal
- "extremely heavy-tailed distributions like the Pareto, all methods exhibit limitations"
  - How likely are we to be in this world?
  - Is income Pareto or log-normal?

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points
- Some framing questions

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points
- Some framing questions
    - Notation is very general, a bit vague $\rightarrow$ give some specific uses

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points
- Some framing questions
  - Notation is very general, a bit vague $\rightarrow$ give some specific uses
  - Embeddings: Take embeddings from LLM/BERT/etc. and then predict $\rightarrow$ factor analysis / Quinn (2004)?

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points
- Some framing questions
    - Notation is very general, a bit vague $\rightarrow$ give some specific uses
    - Embeddings: Take embeddings from LLM/BERT/etc. and then predict $\rightarrow$ factor analysis / Quinn (2004)?
    - Identification in 1D+: Even more complex here b/c of SGD $\rightarrow$ some discussion

Gauthier et al.: Overview

- Auto-encoder to estimate ideal points & effects with big data and many modalities
- Single model to unify analysis & incorporate information from embeddings into ideal points
- Some framing questions
  - Notation is very general, a bit vague $\rightarrow$ give some specific uses
  - Embeddings: Take embeddings from LLM/BERT/etc. and then predict $\rightarrow$ factor analysis / Quinn (2004)?
  - Identification in 1D+: Even more complex here b/c of SGD $\rightarrow$ some discussion
- Simulations do multiple modalities $\rightarrow$ applications are separate?

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?
- Regression and IRT in single model $\rightarrow$ probably undesirable

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?
- Regression and IRT in single model $\rightarrow$ probably undesirable
  - Estimating both together "contaminates" the latent dimension which is (causally?) prior to the outcome

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?
- Regression and IRT in single model $\rightarrow$ probably undesirable
    - Estimating both together “contaminates” the latent dimension which is (causally?) prior to the outcome
    - Especially severe given loss is part of the objective

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?
- Regression and IRT in single model → probably undesirable
  - Estimating both together "contaminates" the latent dimension which is (causally?) prior to the outcome
  - Especially severe given loss is part of the objective
- Critical to calibrate weights on reconstruction/MMD/prediction → guidance for applied users

Gauthier et al.: Method

- What do we gain from using the autoencoder vs., say, stochastic VI/etc. to approximate posterior?
- Regression and IRT in single model $\rightarrow$ probably undesirable
  - Estimating both together "contaminates" the latent dimension which is (causally?) prior to the outcome
  - Especially severe given loss is part of the objective
- Critical to calibrate weights on reconstruction/MMD/prediction $\rightarrow$ guidance for applied users
- More explicit dialogue with existing methods (e.g., Ratkovic!)

Gauthier et al.: Application

- Congressional Data

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress $\rightarrow$ DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?
  - Struggle with face validity in the dimensions:

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?
  - Struggle with face validity in the dimensions:
    - Add a random intercept for sponsor (a la Wordfish) → fix dimension 1

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?
  - Struggle with face validity in the dimensions:
    - Add a random intercept for sponsor (a la Wordfish) → fix dimension 1
    - Some weird dimension 2 (e.g., Americans for Prosperity) → separate general and primary (anti-Trump in primary; anti-Biden in general)

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?
  - Struggle with face validity in the dimensions:
    - Add a random intercept for sponsor (a la Wordfish) → fix dimension 1
    - Some weird dimension 2 (e.g., Americans for Prosperity) → separate general and primary (anti-Trump in primary; anti-Biden in general)
  - Like the examples of phrases

Gauthier et al.: Application

- Congressional Data
  - Separate models for congress → DW-NOMINATE does all congress
    - Temporal variability in words is essential
  - Over-time comparability is...contested... (Bateman et al. 2017)
  - Low correlation with DW-NOMINATE?
  - House results do not make sense
    (Liberalization in both parties after Reconstruction?)
  - Can we use the model to *explain* some of these changes?
- Videos
  - Interesting use case for embeddings
    - What would Quinn (2004) / emIRT give you?
  - Struggle with face validity in the dimensions:
    - Add a random intercept for sponsor (a la Wordfish) → fix dimension 1
    - Some weird dimension 2 (e.g., Americans for Prosperity) → separate general and primary (anti-Trump in primary; anti-Biden in general)
  - Like the examples of phrases
    - "Native elder watches bulldozers destroy burial sites." → conservative?
    - Depictions of police brutality → very moderate?