A Genealogical Application of Statistical Record Linkage for Black Americans in the Antebellum South

SI PROPERTY OF THE PROPERTY OF

Hannah Butler and Andee Kaplan
Department of Statistics, Colorado State University

Overview

- Data on individuals enslaved in the U.S. contain high rates of "alias" records which provide contextual or potentially overlooked information.
- We propose a statistical record linkage model that leverages aliases to link entities with greater success than existing models.
- The model is applied to digitized datasets to attempt to reconstruct more complete histories of these individuals.

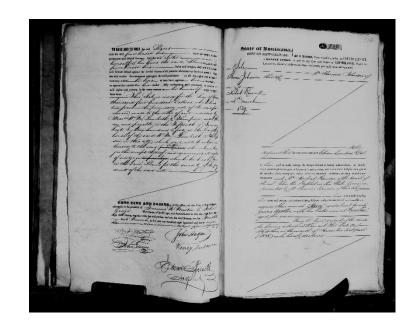
Data

The form that data on enslaved African Americans in the antebellum south took was largely influenced by **oppression**.

Fugitive Advertisements



Notarial Records



Narrative Accounts

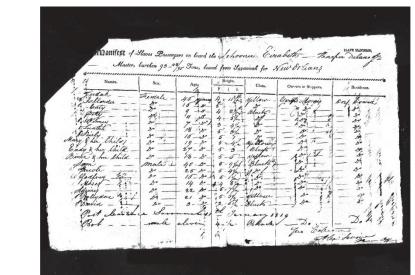
MARY EPPS, ALIAS EMMA BROWN-JOSEPH
AND ROBERT ROBINSON

A SLAVE MOTHER LOSES HER SPEECH AT THE SALE OF HER
CHILD—BOB ESCAPES FROM HIS MASTER, A TRADER,
WITH \$1500 IN NORTH CAROLINA MONEY

Mary fled from Petersburg and the Robinsons from Richmond.
A fugitive slave law-breaking captain by the name of B., who
owned a schooner, and would bring any kind of freight that would
pay the most, was the conductor in this instance. Quite a number
of passengers at different times availed themselves of his accommodations and thus succeeded in reaching Canada.

His risk was very great. On this account he claimed, as did
certain others, that it was no more than fair to charge for his services—indeed he did not profess to bring persons for nothing,
except in rare instances. In this matter the Committee did not
feel disposed to interfere directly in any way, further than to
suggest that whatever understanding was agreed upon by the
parties themselves should be faithfully adhered to.

Ship Manifests



Goal: Develop an effective statistical record linkage model for use in African American genealogical research through the antebellum era.

Alias Records

Alias: The occurrence of one or more duplications of an entity within a file, not due to error, but rather due to a known alternative piece of information.

Entity	Handle	Name	Account	Business
User 1	@colorwired	Abby Joy	Personal	NA
	@mysky.ceramics	My Sky Ceramics	Business	Artist
User 2	@photogeanic	Emma Jean	Personal	NA
	@leafy_jean	E. Jean	Business	Grocery Store
	@oscarbutlerphoto	Oscar	Personal	NA
User 3	@oscarbutlermusic	Oscar Butler	Personal	NA
	@gettingfitwithoscar	Oscar Butler	Personal	NA

What is Record Linkage?

Author	Title	Publ. Date	Author	Title
ephen Jay Gould	The Mismeasure of Man	1996	Meredith Broussard	More than a Glitch: Confronting Race, Gender, and Ability in Tech
ın Hacking	The Taming of Chance	1990	lan Hackng	Taming of Chance
uha Benjamin	Race After Technology	2019	Theodore M. Porter	The Rise of Statistical Thinking
Theodore Porter	Trust in Numbers	2020	Michael E. Staub	The Mismeasure of Minds

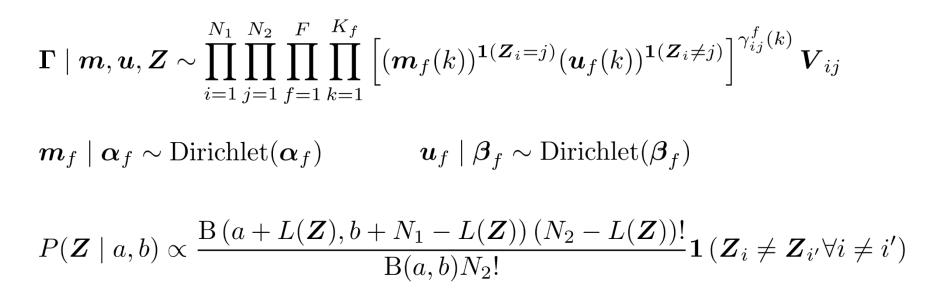
Record linkage encompasses a set of methods used to identify records belonging to a common entity.

- **Deterministic record linkage** uses rules to match field values. A unique identifier, such as an ISBN, makes this a straightforward task.
- **Statistical record linkage** is a statistical model used to estimate matches when no unique identifier is available. This approach offers more flexibility when data is error-prone, and often allows for uncertainty quantification.

Methods

- Specify a **Bayesian bipartite record linkage model** (BRLwA) that accommodates the presence of alias records.
- Choose a **linkage prior** to increase the prior probability that entities with multiple aliases have a greater probability of being linked.

Full Model



- Fit model using **MCMC** to generate posterior samples of the linkage vector and agreement probabilities.
- Use the posterior samples to **estimate links** across data files.

 Γ Comparison data

<----

No Link

- m_f Agreement probabilities (linked)
- $oldsymbol{u}_f$ Agreement probabilities (unlinked)

File A

File B

·----

No Link

- **Z** Linkage vector
- 1. Initialize agreement probabilities and links
- For each iteration, t = 1, . . ., T:
 Sample links, given data and current agreem
- Sample links, given data and current agreement probabilities
 Sample agreement probabilities, given data and current
- links
- 3. Use samples from above to estimate Z

Application

Fugitive Ads X Narrative Accounts

- P(Alias) ~ 0.23; 1,076 of 4,584 entities.
- estimated ~ 0.1% overlap; 5 links by BLRwA.
- Validation not available:
- 1 link manually confirmed.

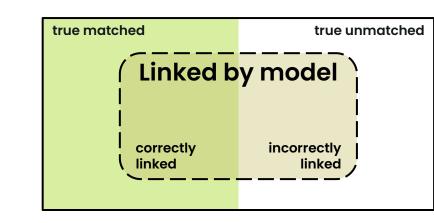
To-Do: Ship Manifests X Notarial Records

- P(Alias) ~ 0.68; 24,187 of 35,442 entities.
- Validation available:
- 0.16% overlap; 58 known matches.

Edgar. Estimated born in 1822. Enslaved by C dward Lewis, alias William Brady. Estimated H Gay in Franklin County, NC. Runaway slave born in 1821. Enslaved in Franklin County, NC by Carter Gay. Attempted to escape advertisement published in 1857. f S150 REWARD. -- RANAWAY FROM the subscriber on the 7th of November, negro slaves, EDGAR and MATT. Edgar is about 35 years old, 6 feet high, of dark brown complexion, almost black, broad shoulders, high cheek bones, long face, and stoops a little in the shoulders. He was raised either in Norfolk or Gloucester County, Va. I bought him in Richmond, July, 1856. The bill of sale was signed by W. Y. Miliner for Jas. A. Bilisoly, ndministrator of G. W. Chambers, dec'd. He told one of go to a free State. As he can read and write it is very probable he has provided himself with some kind of papers, and is making his way to a free State. I will give \$100 for his apprehension and confinement.—Matt. left my premises in July, 1856, is 25 years old, of brown complexion, 5 feet 7 or 8 inches high, high cheek bones, small feet for a negro, wears his hair long, keeps it well combed, and will weigh about 450 pour is. I have heard several times of his being in the upper part of Franklin and Granville counties, sometimes passing himself off under the assumed name of Dunse. I will pay \$50 for his apprehension and confinement so that I get him. of Gay, he and his wife were both sold; the wife one day and he the next. and thirty-five dollars; thus they were sold and resold as a matter of specuour family. Remember us to Mrs. Jackson and Miss Julia. I hope to meet you all so that I get him. My Post Office is Louisburg, Franklin county, N. C. C. H. GAY. SPRINGFIELD ACADEMY, Still, p426-427 Freedom on the Move,

Simulation Results

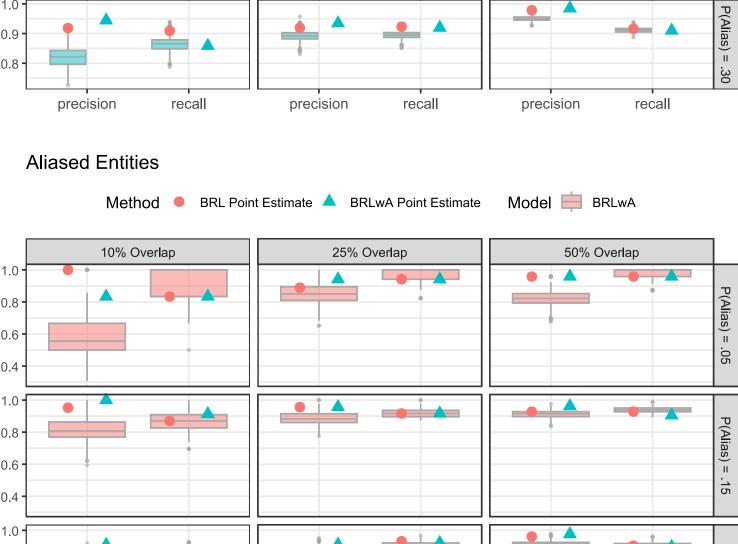
- Tested BRLwA on simulated data against Bayesian model (BRL), with post-hoc conflict resolution.
- BRLwA point estimates of Z show comparable precision & recall to BRL (with correction).
- BRLwA maintains posterior distribution, allowing for uncertainty quantification.

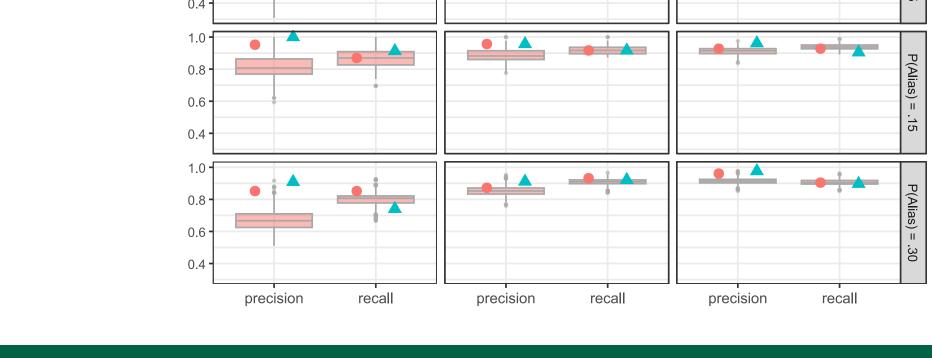


Measuring Performance

Precision: (correctly linked) / (all linked)
Recall: (correctly linked) / (true matched)







References

- Fellegi, Ivan P., and Alan B. Sunter. "A Theory for Record Linkage." Journal of the American Statistical Association 64, no. 328 (1969): 1183–1210. https://doi.org/10.2307/2286061.
- Freedom on the Move. "Freedom on the Move." Accessed October 30, 2024.
- Nick Sacco. "William Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 https://docs.google.com/spreadsheets/d/e/2PACX-lvTpXEcSlrao3f-ZHqeOApPoBXju6-
 **Comparison of the standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed November 5, 2024.
 **The standard Still Underground Railroad Data.Xlsx." Accessed
- Xeg902ujGjuDqkE_cxd0MGl8P-DFv4SMPqbg/pubhtml?usp=embed_facebook.
 Kinfolkology. "Kinfolkology." Accessed October 30, 2024. https://www.kinfolkology.org.
 Sadinle, Mauricio. "Bayesian Estimation of Bipartite Matchings for Record Linkage." Journal of the
- American Statistical Association 112, no. 518 (April 3, 2017): 600–612.

 https://doi.org/10.1080/01621459.2016.1148612.

 Still, William. The Underground Railroad: A Record of Facts, Authentic Narratives, Letters &c.,
- Narrating the Hardships, Hair-Breadth Escapes and Death Struggles of the Slaves in Their Efforts for Freedom. Rev. ed. Philadelphia, Pa., Cincinnati, Ohio [etc.]: People's publishing company, 1879.
- https://www.loc.gov/resource/rbc0001.2019gen24984/
 Still, William. "The Underground Railroad., Rev. Ed." Text.

https://doi.org/10.1080/0144039X.2019.166050

- https://www.gutenberg.org/files/15263/15263-h/15263-h.htm. Accessed November 13, 2024.
- Williams, Jennie K. "Trouble the Water: The Baltimore to New Orleans Coastwise Slave Trade, 1820-1860." Slavery & Abolition 41, no. 2 (April 2, 2020): 275–303.

Acknowledgements

This work is supported in part by the National Science Foundation under Grant No. SES-2338428 and DMS-2330089. The ideas in this work are representative of the authors and not of the NSF.