Curriculum Vitae
December, 2021

1. Personal Details

First name: Nir
Last name: Billfeld
Permanent Address: The Weizmann Institute of Science, Department of Computer Science and Applied Mathematics.
Telephone number: +972 528707194
Email: nir.billfeld@weizmann.ac.il
Skype: nirbillfeld@gmail.com
Homepage: https://www.nirbillfeld.com

2. Higher Education

Institution	Department	Field of study	Degree	Period of study
University of Haifa	Statistics		B.A.	2002-2006
University of Haifa	Economics		B.A.	2003-2006
Tel-Aviv University	Economics	Labor Economics (the panel data - HRS survey)	M.A.	2006-2011
University of Haifa	Economics	Econometrics	P.h.D.	2014-2019
The Weizmann Institute of Science	Computer Science and Applied Mathematics	Econometircs – Causal Inference with GANs Using Pytorch with multiple GPUs.	Post doctorate fellowship	2020-2022

PhD Dissertation

Truncation Bias

Supervised by Prof. Moshe Kim.

3. List of Publications

Billfeld, N. and Kim, M. (2019). Semiparametric Wavelet-based JPEG IV Estimator for Endogenously Truncated Data. IEEE Access, 7, 99602 - 99621.
DOI (identifier) 10.1109/ACCESS.2019.2929571

Available at IEEE: https://ieeexplore.ieee.org/document/8765569

Job Market Paper

ABSTRACT

A new and an enriched JPEG algorithm is provided for identifying redundancies in a sequence of irregular noisy data points which also accommodates a reference-free criterion function. Our main contribution is by formulating analytically (instead of approximating) the inverse of the transpose of JPEG-wavelet transform without involving matrices which are computationally cumbersome. The algorithm is suitable for the widely-spread situations where the original data distribution is unobservable such as in cases where there is deficient representation of the entire population in the training data (in machine learning) and thus the covariate shift assumption is violated. The proposed estimator corrects for both biases, the one generated by endogenous truncation and the one generated by endogenous covariates. Results from utilizing 2,000,000 different distribution functions verify the applicability and high accuracy of our procedure to cases in which the disturbances are neither jointly nor marginally normally distributed.

Billfeld, N., and Kim, M. (2019). Semiparametric Correction for Endogenous Truncation Bias With Vox Populi-Based Participation Decision. IEEE Access, 7, 12114-12132.
DOI (identifier) 10.1109/ACCESS.2018.2888575
Available at IEEE: https://ieeexplore.ieee.org/document/8580454

ABSTRACT

We synthesize the knowledge present in various scientific disciplines for the development of semiparametric endogenous truncation-proof algorithm, correcting for truncation bias due to endogenous self-selection. This synthesis enriches the algorithm's accuracy, efficiency and applicability. Improving upon the covariate shift assumption, data are intrinsically affected and largely generated by their own behavior (cognition). Refining the concept of Vox Populi (Wisdom of Crowd) allows data points to sort themselves out depending on their estimated latent reference group opinion space. Monte Carlo simulations, based on 2,000,000 different distribution functions, practically generating 100 million realizations, attest to a very high accuracy of our model.

Billfeld, N. and Kim, M. (2018). Semiparametric Maximum Likelihood Sieve Estimator for Correction of Endogenous Truncation Bias.

ERN Econometric & Statistical Methods - General eJournal, Vol. 12 No. 1, Jan 4, 2019

SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286553

ABSTRACT

Semiparametric correction for a sample selection bias in the presence of endogenous truncation is known to be much more di cult in the case of a binary selection variable than in the case of a continuous selection variable. This paper proposes a simple bandwidth-free semiparametric methodology to correct for a self-selection bias in a truncated sample, without any prior knowledge of the marginal density functions of the selection model's random disturbances. Each of the unknown marginal density functions is estimated using Sieve estimator, utilizing Hermite polynomials as basis functions. The aforementioned procedure is appropriate for both binary and continuous selection variables cases under the covariate shift assumption. We consider a double hurdle model, which is a combination of two selection rules. The first is propagated by a truncation in the dependent variable of the substantive equation. The second is propagated by endogenous self-selection. The suggested correction procedure produces estimates that are of high accuracy and consistent based on Monte Carlo simulations. The random disturbances are not restricted to being symmetric and their

marginal distribution functions are unknown. Thus, in the data generation process we verify the applicability of our procedure to cases in which the disturbances are neither jointly nor marginally normally distributed. These disturbances are constructed as realizations of non-symmetric distribution functions.

Billfeld, N. and Kim, M. (2019). Endogenous Latent Masking: Neutralizing the Bias.

Masking in the context of partial observability is a widely-spread phenomenon in the areas of robotics, pattern recognition, signal recovery and image processing, to mention a few. An important and overlooked issue is that latent covariates (unobservables), represented by a joint distribution of random disturbances, very often generate comovements between the process underlying the masking and the distribution of the outcome variable, even when holding the observable covariates constant. The endogeneity embedded in the masking variable further aggravates the discrepancy bias between the observed and the real world data. Simply put, endogeneity in the masked data reflects the violation of the prevailing i.i.d zero mean white noise, distorting thus biasing parameter estimates. Consequently, various image denoising filters relying on these paramaters are no longer valid, such as, e.g, JPEG. Bridging the gap between dimensionality reduction and flexibility reflected in nonparametric distributions, we offer a semiparametric identification and estimation strategy floating the unmasked (original) joint distribution of these latent covariates, enabling the removal of their associated bias from the desired parameter estimates. The novelty of the suggested model is in neutralizing the bias generated by the discrepancy between reality and masking rather than targeting the impossible effort to mimic reality.

Billfeld, N. and Kim, M. (2018). Nonseparability without Monotonicity: The Couterfactual Distribution for Causal Inference.

SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3343438

ABSTRACT

Nonparametric identification strategy is employed to capture causal relationships without imposing any variant of monotonicity existing in the nonseparable nonlinear error model literature. This is important as when monotonicity is applied to the instrumental variables it limits their availability and when applied to the unobservables it can hardly be justified in the non-scalar case. Moreover, in cases where monotonicity is not satisfied the monotonicity-based estimators might be severely biased as shown in comparative Monte Carlo simulation. The key idea in the proposed identification and estimation strategy is to uncover the counterfactual distribution of the dependent variable, which is not directly observed in the data. We o er a two-step M-Estimator based on a resolution-dependent reproducing symmetric kernel density estimator rather than on the bandwidth-dependent classical kernel and thus, less sensitive to bandwidth choice. Additionally, the average marginal e ect of the endogenous covariate on the outcome variable is identified directly from the noisy data which precludes the need to employ additional estimation steps thereby avoiding potential error accumulation. Asymptotic properties of the counterfactual M-Estimator are established.

Billfeld, N. and Kim, M. (2018). Semiparametric Fourier-dependent Sieve IV estimator (SPIV) for truncated data. (Submitted)

SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3308012

ABSTRACT

Instrumental variables are intended to correct for misspecifications largely stemming from endogeneity problems or omission of relevant important covariates correlated with some of the other included covariates. The validity of the IV estimator relies on the orthogonality with respect to the random disturbance. However, in cases of endogenously truncated data as well as in other instances (e.g, censored data) which is very frequently the nature of data used in empirical research, there exists severe contamination in the disturbance due to the endogenous selection process. The endogenous selection process generates a co-movement between the IV and the disturbance which is related to the variation in the selection equation's covariates. This contamination propagates additional bias introduced into the parameter estimates of the various covariates. Consequently, not only that the conventional IV does not solve the problem it is intended to but rather introduces additional bias into the parameter estimates of the various covariates of the substantive equation. Our empirical implementation shows that even under mild correlation between the random disturbances, the resulting bias in the estimated parameter of the endogenous covariate in the substantive equation can amount to almost 10 times the true parameter value for 500 observations and can amount to 5 times the true parameter value in a sample of 10,000 observations. We offer a semi-parametric Fourier-dependent Sieve IV (SPIV) estimator correcting for both truncation as well as endogeneity biases. The proposed estimator removes the hurdle which prevents orthogonality under truncation or other misspecifications. Using Monte Carlo simulations attest to the high accuracy of our offered semi-parametric Sieve IV estimator as expressed by the $\sqrt{n}$ consistency. These results have been verified by utilizing 2,000,000 different distribution functions, practically generating 100 millions realizations to generate the various data sets.

4. Works in progress

Billfeld, N. and Kim M., Cancer Detection by Multi-Agent Swarms intelligence.

Billfeld, N. and Kim M., Causal Non-Eergodic-Graphs (NEG): Latent Contextual Neighbors.

Billfeld, N. and Kim M., Coplula and Nested Mean models.

Billfeld, N. and Kim, M. Nonlinear Diff-In-Diff in Nonseparable Error Models.

Billfeld, N. and Kim, M. Swarms Participation Decision in Truncated Data.

Billfeld, N. and Kim, M. Nonseparability without Monotonicity: the Counterfactual Distribution Estimator for Causal Inference.

5. Scholarships, Awards, Research Grants or participation in Research, Grants etc.

Award \ Scholarship	Department\Faculty	Date
Aaron Guttroyt Foundation Scholarship	Statistic	3.2014
Certificate of Excellence For PhD students	The Social Science Faculty	6.2017

6.Work Experience

Institution	Location	Job description	Years
Applied Materials	Rehovot	Deep learning algorithm developer	12/2021-
Weizmann Institute of Sciene	Rehovot	Machine learning and Econometrics algorithm developer for causal inference	01/2020-12/2021
Ort Braude College of Engineering \ Western Galil College	**Karmiel\Ako**	Deployed as Teaching Assistant and Lecturer to teach Econometrics, Economics and Statistics to BSc students.	2011-2019
Israel Railways	**Tel-Aviv**	Data Analyst	2010-2011
Applied Economics	**Tel-Aviv**	Economic Data Analyst	2009-2010
Foerder Institute for Economic Research	Tel-Aviv (University)	Statistical Models Programmer	2008 – 2009

7. Teaching

Title	Year	Type	Level
Econometrics	2011-2016	Lecturer	B.SC.
Economic growth models	2012-2016	Lecturer	B.SC.
Introduction to macro Economics	2012-2015	Lecturer	B.SC.
Introduction to micro Economics	2012-2015	Lecturer	B.SC.
Introduction to macro Economics	2012-2016	Teaching Assistant	B.A.
Introduction to micro Economics	2012-2016	Teaching Assistant	B.A.
Introduction to probability and statistics	2013-2018	Teaching Assistant	B.A.
	2017-2018	Teaching Assistant	B.A.

Talks

Date	Seminar	Place	Topic
November, 2017	Economics Department Seminar	Bar-Ilan University Ramat-Gan, Israel	SEMIPARAMETRIC FOURIER-DEPENDENT SIEVE IV ESTIMATOR
Januray, 2019	Machine Learning and Statistics Seminar	THE WEIZMANN INSTITUTE OF SCIENCE Rehovot, Israel	Semiparametric Wavelet-based JPEG IV Estimator for Endogenously Truncated Data
Januray, 2019	The Israel Statistical Association Doctorants Seminar	Tel-Aviv University Tel-Aviv, Israel	Semiparametric Wavelet-based JPEG IV Estimator for Endogenously Truncated Data

8. Referee

IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE ACCESS.

9. Languages:

- **Hebrew (full proficiency), English (full proficiency).**

- **Proficient in several statistical programs, such as R, STATA, MATLAB, Pythone. Object Oriented Programming in C++.**