

# Lars Vilhuber - Publications

Lars Vilhuber

December 18, 2020

## Contents

<b>1</b>	<b>Referreed Publications</b>	<b>1</b>
<b>2</b>	<b>Proceedings</b>	<b>14</b>
<b>3</b>	<b>Book chapters</b>	<b>22</b>
<b>4</b>	<b>Working papers and other documents</b>	<b>24</b>
<b>5</b>	<b>Online resources</b>	<b>49</b>

## 1 Referreed Publications

**Alam et al.: Applying Data Synthesis for Longitudinal Business Data across Three Countries** **alam2020**

---

M. Jahangir Alam, Benoit Dostie, Jörg Drechsler, and Lars Vilhuber. “Applying Data Synthesis for Longitudinal Business Data across Three Countries”. In: *Statistics in Transition New Series* 21.4 (2020), pp. 212–236. DOI: [10.21307/stattrans-2020-039](https://doi.org/10.21307/stattrans-2020-039).

File: [Alametal\\_2020\\_Applyingdatasynthesisforlongitudinalbusinessdataacrossthreecountries.pdf](#) : /home / vilhuber / Zotero / storage / 35RKH8NM / Alametal \_ 2020 \_ Applyingdatasynthesisforlongitudinalbusinessdataacrossthreecountries . pdf:application/pdf.

**Foote et al.: Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings**

**FooteKutzbachVilhuber-submitted**

---

Andrew Foote, Mark J. Kutzbach, and Lars Vilhuber. “Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings”. In: *Applied Economics* accepted (2020).

Abstract: This paper evaluates the use of commuting zones as a local labor market definition. We revisit Tolbert and Sizer (1996) and demonstrate the sensitivity of definitions to two features of the methodology. We show how these features impact empirical estimates using a well-known application of commuting zones. We conclude with advice to researchers using commuting zones on how to demonstrate the robustness of empirical findings to uncertainty in definitions.

**K. L. McKinney et al.: Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On-TheMap**

**McKinneyEtAl:accepted:2017**

---

Kevin L. McKinney, Andrew S. Green, John M. Abowd, and Lars Vilhuber. “Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in OnTheMap”. In: *Journal of Survey Statistics and Methodology* accepted (2020).

Abstract: We report results from the first comprehensive total quality evaluation of five major indicators in the U.S. Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) Program Quarterly Workforce Indicators (QWI): total employment, beginning-of-quarter employment, full-quarter employment, total payroll, and average monthly earnings of full-quarter employees. Beginning-of-quarter employment is also the main tabulation variable in the LEHD Origin-Destination Employment Statistics (LODES) workplace reports as displayed in OnTheMap (OTM). The evaluation is conducted by generating multiple threads of the edit and imputation models used in the LEHD Infrastructure File System. These threads conform to the Rubin (1987) multiple imputation model, with each thread or implicate being the output of formal probability models that address coverage, edit, and

imputation errors. Design-based sampling variability and finite population corrections are also included in the evaluation. We derive special formulas for the Rubin total variability and its components that are consistent with the disclosure avoidance system used for QWI and LODES/OTM workplace reports. These formulas allow us to publish the complete set of detailed total quality measures for QWI and LODES. The analysis reveals that the five publication variables under study are estimated very accurately for tabulations involving at least 10 jobs. Tabulations involving three to nine jobs have quality in the range generally deemed acceptable. Tabulations involving zero, one or two jobs, which are generally suppressed in the QWI and synthesized in LODES, have substantial total variability but their publication in LODES allows the formation of larger custom aggregations, which will in general have the accuracy estimated for tabulations in the QWI based on a similar number of workers.

### **Vilhuber: Reproducibility and Replicability in Economics** **Vilhuber2020Reproducibility**

---

Lars Vilhuber. “Reproducibility and Replicability in Economics”. In: *Harvard Data Science Review* 2.4 (2020). DOI: [10.1162/99608f92.4f6b9e67](https://doi.org/10.1162/99608f92.4f6b9e67).

### **Vilhuber et al.: Report by the AEA Data Editor**    **ReportDE2020**

---

Lars Vilhuber, James Turrito, and Keesler Welch. “Report by the AEA Data Editor”. In: *AEA Papers and Proceedings* 110 (2020), pp. 764–75. DOI: [10.1257/pandp.110.764](https://doi.org/10.1257/pandp.110.764).

### **John M. Abowd et al.: Why the Economics Profession Must Actively Participate in the Privacy Protection Debate**    **AEAPP2019**

---

John M. Abowd, Ian M. Schmutte, William N. Sexton, and Lars Vilhuber. “Why the Economics Profession Must Actively Participate in the Privacy Protection Debate”. In: *AEA Papers and Proceedings* 109 (2019), pp. 397–402. DOI: [10.1257/pandp.20191106](https://doi.org/10.1257/pandp.20191106).

Abstract: When Google or the U.S. Census Bureau publish detailed statistics on browsing habits or neighborhood characteristics, some privacy is lost for everybody while supplying public information. To date, economists have not focused on the privacy loss inherent in data publication. In their stead, these

issues have been advanced almost exclusively by computer scientists who are primarily interested in technical problems associated with protecting privacy. Economists should join the discussion, first, to determine where to balance privacy protection against data quality; a social choice problem. Furthermore, economists must ensure new privacy models preserve the validity of public data for economic research.

---

**Vilhuber: Report by the AEA Data Editor** **ReportDE2019**

---

Lars Vilhuber. “Report by the AEA Data Editor”. In: *AEA Papers and Proceedings* 109 (2019), pp. 718–29. DOI: [10.1257/pandp.109.718](https://doi.org/10.1257/pandp.109.718).

---

**Slavković et al.: Remembering Stephen Fienberg** **Slavkovic2018**

---

Aleksandra Slavković and Lars Vilhuber. “Remembering Stephen Fienberg”. In: *Journal of Privacy and Confidentiality* 8.1 (2018). DOI: [10.29012/jpc.685](https://doi.org/10.29012/jpc.685).

---

**Vilhuber: Relaunching the Journal of Privacy and Confidentiality**  
**Vilhuber2018**

---

Lars Vilhuber. “Relaunching the Journal of Privacy and Confidentiality”. In: *Journal of Privacy and Confidentiality* 8.1 (2018). DOI: [10.29012/jpc.706](https://doi.org/10.29012/jpc.706).

---

**Weinberg et al.: Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Improve the U.S. Statistical System?**  
**ncrn-summary**

---

Daniel H. Weinberg et al. “Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Improve the U.S. Statistical System?” In: *Journal of Survey Statistics and Methodology* (2018). DOI: [10.1093/jssam/smy023](https://doi.org/10.1093/jssam/smy023).

Abstract: The National Science Foundation-Census Bureau Research Network (NCRN) was established in 2011 to create interdisciplinary research nodes on methodological questions of interest and significance to the broader research community and to the Federal Statistical System (FSS), particularly the Census Bureau. The activities to date have covered both fundamental and applied statistical research and have focused at least in part on the training of current and future generations of researchers in skills of relevance

to surveys and alternative measurement of economic units, households, and persons. This paper discusses some of the key research findings of the eight nodes, organized into six topics: (1) Improving census and survey data collection methods; (2) Using alternative sources of data; (3) Protecting privacy and confidentiality by improving disclosure avoidance; (4) Using spatial and spatio-temporal statistical modeling to improve estimates; (5) Assessing data cost and quality tradeoffs; and (6) Combining information from multiple sources. It also reports on collaborations across nodes and with federal agencies, new software developed, and educational activities and outcomes. The paper concludes with an evaluation of the ability of the FSS to apply the NCRN's research outcomes and suggests some next steps, as well as the implications of this research-network model for future federal government renewal initiatives.

**Cloutier et al.: Understanding the effect of procedural justice on psychological distress** **CloutierVilhuber2017**

---

Julie Cloutier, Lars Vilhuber, Denis Harrisson, and Vanessa Béland-Ouellette. "Understanding the effect of procedural justice on psychological distress". In: *International Journal of Stress Management* Advance online (2017). DOI: [10.1037/str0000065](https://doi.org/10.1037/str0000065).

**Abstract:** Studies on the effect of procedural justice on psychological distress present conflicting results. Drawing on instrumental and relational perspectives of justice, we test the hypothesis that the perception of procedural justice influences the level of workers' psychological distress. Using a number of validated instruments to collected data from 659 workers in three call centers, we use OLS regressions and Hayes' PROCESS tool to show that the perception of procedural justice has a direct, unique, and independent effect on psychological distress. The perception of procedural justice has no instrumental role, the key mechanism being the relational role, suggesting that perceived injustice influences psychological distress because it threatens self-esteem. Distributive justice perceptions (recognition, promotions, job security) are not associated with psychological distress, calling into question Siegrist's model. Our findings suggest that perceived procedural justice provides workers better evidence of the extent to which they are valued and appreciated members of their organizations than do perceptions of distributive justice. The results highlight the greater need for workers to be valued

and appreciated for who they are (consideration and esteem), rather than for what they do for their organization (distributive justice of rewards).

**Vilhuber et al.: Making Confidential Data Part of Reproducible Research**  
**chance:2017**

---

Lars Vilhuber and Carl Lagoze. “Making Confidential Data Part of Reproducible Research”. In: *Chance* (2017). URL: <http://chance.amstat.org/2017/09/reproducible-research/>.

**Miranda et al.: Using partially synthetic microdata to protect sensitive cells in business statistics**

**MirandaVilhuber:Using:SJIAOS:2016**

---

Javier Miranda and Lars Vilhuber. “Using partially synthetic microdata to protect sensitive cells in business statistics”. In: *Statistical Journal of the International Association for Official Statistics* 32.1 (2016), pp. 69–80. DOI: [10.3233/SJI-160963](https://doi.org/10.3233/SJI-160963).

Abstract: We describe and analyze a method that blends records from both observed and synthetic microdata into public-use tabulations on establishment statistics. The resulting tables use synthetic data only in potentially sensitive cells. We describe different algorithms, and present preliminary results when applied to the Census Bureau’s Business Dynamics Statistics and Synthetic Longitudinal Business Database, highlighting accuracy and protection afforded by the method when compared to existing public-use tabulations (with suppressions).

**Vilhuber et al.: Synthetic establishment microdata around the world**  
**VilhuberAbowdReiter:Synthetic:SJIAOS:2016**

---

Lars Vilhuber, John M. Abowd, and Jerome P. Reiter. “Synthetic establishment microdata around the world”. In: *Statistical Journal of the International Association for Official Statistics* 32.1 (2016), pp. 65–68. DOI: [10.3233/SJI-160964](https://doi.org/10.3233/SJI-160964).

Abstract: In contrast to the many public-use microdata samples available for individual and household data from many statistical agencies around the world, there are virtually no establishment or firm microdata available. In large part, this difficulty in providing access to business micro data is

due to the skewed and sparse distributions that characterize business data. Synthetic data are simulated data generated from statistical models. We organized sessions at the 2015 World Statistical Congress and the 2015 Joint Statistical Meetings, highlighting work on synthetic establishment micro-data. This overview situates those papers, published in this issue, within the broader literature.

**Jörg Drechsler et al.: A First Step Towards A German SynLBD:  
Constructing A German Longitudinal Business Database**

**SJIAOS-2014b**

---

Jörg Drechsler and Lars Vilhuber. “A First Step Towards A German SynLBD: Constructing A German Longitudinal Business Database”. In: *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics* 30 (2014). DOI: [10.3233/SJI-140812](https://doi.org/10.3233/SJI-140812).

Abstract: One major criticism against the use of synthetic data has been that the efforts necessary to generate useful synthetic data are so intense that many statistical agencies cannot afford them. We argue many lessons in this evolving field have been learned in the early years of synthetic data generation, and can be used in the development of new synthetic data products, considerably reducing the required investments. The final goal of the project described in this paper will be to evaluate whether synthetic data algorithms developed in the U.S. to generate a synthetic version of the Longitudinal Business Database (LBD) can easily be transferred to generate a similar data product for other countries. We construct a German data product with information comparable to the LBD - the German Longitudinal Business Database (GLBD) - that is generated from different administrative sources at the Institute for Employment Research, Germany. In a future step, the algorithms developed for the synthesis of the LBD will be applied to the GLBD. Extensive evaluations will illustrate whether the algorithms provide useful synthetic data without further adjustment. The ultimate goal of the project is to provide access to multiple synthetic datasets similar to the SynLBD at Cornell to enable comparative studies between countries. The Synthetic GLBD is a first step towards that goal.

**Miranda et al.: Looking Back On Three Years Of Using The Synthetic LBD Beta**  
**SJIAOS-2014a**

---

Javier Miranda and Lars Vilhuber. “Looking Back On Three Years Of Using The Synthetic LBD Beta”. In: *Statistical Journal of the IAOS: Journal of the International Association for Official Statistics* 30 (2014). DOI: [10.3233/SJI-140811](https://doi.org/10.3233/SJI-140811).

Abstract: Distributions of business data are typically much more skewed than those for household or individual data and public knowledge of the underlying units is greater. As a results, national statistical offices (NSOs) rarely release establishment or firm-level business microdata due to the risk to respondent confidentiality. One potential approach for overcoming these risks is to release synthetic data where the establishment data are simulated from statistical models designed to mimic the distributions of the real underlying microdata. The US Census Bureau’s Center for Economic Studies in collaboration with Duke University, the National Institute of Statistical Sciences, and Cornell University made available a synthetic public use file for the Longitudinal Business Database (LBD) comprising more than 20 million records for all business establishment with paid employees dating back to 1976. The resulting product, dubbed the SynLBD, was released in 2010 and is the first-ever comprehensive business microdata set publicly released in the United States including data on establishments employment and payroll, birth and death years, and industrial classification. This paper documents the scope of projects that have requested and used the SynLBD.

**John M. Abowd et al.: Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation**

**AbowdSchneiderVilhuber2013**

---

John M. Abowd, Matthew J. Schneider, and Lars Vilhuber. “Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation”. In: *Journal of Privacy and Confidentiality* 5.1 (2013). Article 4. URL: <https://doi.org/10.29012/jpc.v5i1.627>.

Abstract: We consider a particular maximum likelihood estimator (MLE) and a computationally intensive Bayesian method for differentially private estimation of the linear mixed-effects model (LMM) with normal random errors. The LMM is important because it is used in small-area estimation



and detailed industry tabulations that present significant challenges for confidentiality protection of the underlying data. The differentially private MLE performs well compared to the regular MLE, and deteriorates as the protection increases for a problem in which the small-area variation is at the county level. More dimensions of random effects are needed to adequately represent the time dimension of the data, and for these cases the differentially private MLE cannot be computed. The direct Bayesian approach for the same model uses an informative, reasonably diffuse prior to compute the posterior predictive distribution for the random effects. The empirical differential privacy of this approach is estimated by direct computation of the relevant odds ratios after deleting influential observations according to various criteria.

File: [AbowdSchneiderVilhuber2013.pdf:A/AbowdSchneiderVilhuber2013.pdf:PDF](#).

### **Jörg Drechsler et al.: Replicating the Synthetic LBD with German Establishment Data** **ISI2013-3**

---

Jörg Drechsler and Lars Vilhuber. “Replicating the Synthetic LBD with German Establishment Data”. In: *Proceedings 59th ISI World Statistics Congress, 25-30 August 2013, Hong Kong (Session STS062)* (2013), pp. 2291–2296. URL: <http://2013.isiproceedings.org> (visited on 03/24/2014).

### **Lagoze et al.: Data Management of Confidential Data** **DBLP:journals/ijdc/LagozeBWAV13**

---

Carl Lagoze, William C. Block, Jeremy Williams, John M. Abowd, and Lars Vilhuber. “Data Management of Confidential Data”. In: *International Journal of Digital Curation* 8.1 (2013), pp. 265–278. DOI: [10.2218/ijdc.v8i1.259](https://doi.org/10.2218/ijdc.v8i1.259).

**Abstract:** Social science researchers increasingly make use of data that is confidential because it contains linkages to the identities of people, corporations, etc. The value of this data lies in the ability to join the identifiable entities with external data such as genome data, geospatial information, and the like. However, the confidentiality of this data is a barrier to its utility and curation, making it difficult to fulfill US federal data management mandates and interfering with basic scholarly practices such as validation and reuse of existing results. We describe the complexity of the relationships among

data that span a public and private divide. We then describe our work on the CED2AR prototype, a first step in providing researchers with a tool that spans this divide and makes it possible for them to search, access, and cite that data.

**John M. Abowd et al.: Did the Housing Price Bubble Clobber Local Labor Market Job and Worker Flows When It Burst?**

**AbowdVilhuber2012**

---

John M. Abowd and Lars Vilhuber. “Did the Housing Price Bubble Clobber Local Labor Market Job and Worker Flows When It Burst?” In: *American Economic Review* 102.3 (2012), pp. 589–93. DOI: [10.1257/aer.102.3.589](https://doi.org/10.1257/aer.102.3.589).

Abstract: We use the Census Bureau’s Quarterly Workforce Indicators and the Federal Housing Finance Agency’s House Price Indices to study the effects of the housing price bubble on local labor markets. We show that the 35 MSAs in the top decile of the house price boom were most severely impacted. Their stable job employment fell much more than the national average. Their real wage rates did not fall as fast as the national average. Accessions fell much faster than average while separations were constant. Job creations fell substantially while destructions rose slightly.

**John M. Abowd et al.: Did the Housing Price Bubble Clobber Local Labor Market Job and Worker Flows When It Burst? - Online Appendix**

**aea-pp-abowd-vilhuber-online-appendix**

---

John M. Abowd and Lars Vilhuber. “Did the Housing Price Bubble Clobber Local Labor Market Job and Worker Flows When It Burst? - Online Appendix”. In: *American Economic Review* 102.3 (2012), pp. 589–93. DOI: [10.1257/aer.102.3.589](https://doi.org/10.1257/aer.102.3.589).

Abstract: This is the appendix to Abowd and Vilhuber (2012).

**John M. Abowd et al.: National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail**

**AbowdVilhuber2010**

---

John M. Abowd and Lars Vilhuber. “National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demo-

graphic and Industry Detail”. In: *Journal of Econometrics* 161 (2011), pp. 82–99. DOI: [10.1016/j.jeconom.2010.09.008](https://doi.org/10.1016/j.jeconom.2010.09.008).

Abstract: The Quarterly Workforce Indicators (QWI) are local labor market data produced and released every quarter by the United States Census Bureau. Unlike any other local labor market series produced in the US or the rest of the world, QWI measure employment flows for workers (accession and separations), jobs (creations and destructions) and earnings for demographic subgroups (age and gender), economic industry (NAICS industry groups), detailed geography (block (experimental), county, Core-Based Statistical Area, and Workforce Investment Area), and ownership (private, all) with fully interacted publication tables. The current QWI data cover 47 states, about 98% of the private workforce in those states, and about 92% of all private employment in the entire economy. State participation is sufficiently extensive to permit us to present the first national estimates constructed from these data. We focus on worker, job, and excess (churning) reallocation rates, rather than on levels of the basic variables. This permits a comparison to existing series from the Job Openings and Labor Turnover Survey and the Business Employment Dynamics Series from the Bureau of Labor Statistics (BLS). The national estimates from the QWI are an important enhancement to existing series because they include demographic and industry detail for both worker and job flow data compiled from underlying micro-data that have been integrated at the job and establishment levels by the Longitudinal Employer-Household Dynamics Program at the Census Bureau. The estimates presented herein were compiled exclusively from public-use data series and are available for download.

File: [AbowdVilhuber2010.pdf:A/AbowdVilhuber2010.pdf:PDF](#).

### **John M. Abowd et al.: Science, Confidentiality, and the Public Interest** **Chance2011**

---

John M. Abowd and Lars Vilhuber. “Science, Confidentiality, and the Public Interest”. In: *Chance* 24.3 (2011), pp. 58–62. DOI: [10.1080/09332480.2011.10739876](https://doi.org/10.1080/09332480.2011.10739876).

**Cloutier et al.: Procedural justice criteria in salary determination**  
**CloutierVilhuber2008**

---

Julie Cloutier and Lars Vilhuber. “Procedural justice criteria in salary determination”. In: *Journal of Managerial Psychology* 23.6 (2008), pp. 713–740. DOI: [10.1108/02683940810894765](https://doi.org/10.1108/02683940810894765).

**Abstract:** Purpose – The purpose of this research is to identify the dimensionality of the procedural justice construct and the criteria used by employees to assess procedural justice, in the context of salary determination. Design/methodology/approach – Based on a survey of 297 Canadian workers, the paper uses confirmatory factor analysis (CFA) to test the dimensionality and the discriminant and convergent validity of our procedural justice construct. Convergent and predictive validity are also tested using hierarchical linear regressions. Findings – The paper shows the multidimensionality of the procedural justice construct: justice of the salary determination process is assessed through the perceived characteristics of allocation procedures, the perceived characteristics of decision-makers, and system transparency. Research limitations/implications – Results could be biased towards acceptance; this is discussed. The results also suggest possible extensions to the study. Practical implications – Knowledge of the justice standards improves the ability of organizations to effectively manage the salary determination process and promote its acceptance among employees. Emphasizes the need to adequately manage the selection, training, and perception of decision makers. Originality/value – The paper identifies the standards of procedural justice for salary determination processes. It contributes to the theoretical literature by providing a new multidimensional conceptualization, which helps to better understand the psychological process underlying the perception of procedural justice. The presence of a dimension associated with decision makers is novel and critical for compensation studies.

**Machanavajjhala et al.: Privacy: Theory meets practice on the map**  
**Ashwin2008**

---

Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. “Privacy: Theory meets practice on the map”. In: *International Conference on Data Engineering (ICDE)* (2008), pp. 277–286. DOI: [10.1109/ICDE.2008.4497436](https://doi.org/10.1109/ICDE.2008.4497436).

**John M. Abowd et al.: The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers** **AbowdVilhuber2005**

---

John M. Abowd and Lars Vilhuber. “The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers”. In: *Journal of Business and Economic Statistics* 23.2 (2005), pp. 133–152. URL: <http://www.jstor.org/stable/27638803>.

File: [AbowdVilhuber2005.pdf:A/AbowdVilhuber2005.pdf:PDF;tp-2002-17.pdf:L/LEHD/tp-2002-17.pdf:PDF](#).

**H. Holzer et al.: Escaping poverty for low-wage workers: The role of employer characteristics and changes** **HolzerLaneVilhuber2004**

---

Harry Holzer, Julia Lane, and Lars Vilhuber. “Escaping poverty for low-wage workers: The role of employer characteristics and changes”. In: *Industrial and Labor Relations Review* 57.4 (2004). DOI: [10.1177/001979390405700405](https://doi.org/10.1177/001979390405700405).

Abstract: Using a unique dataset based on individual Unemployment Insurance wage records for Illinois in the 1990s that are matched to other Census data, the authors analyze the extent to which escape from or entry into low earnings among adult workers was associated with changes in their employers and firm characteristics. The results show considerable mobility into and out of low earnings status, even for adults. They indicate that job changes were an important part of the process by which workers escaped or entered low-wage status, and that changes in employer characteristics help to account for these job changes. Matches between personal and firm characteristics also contributed to observed earnings outcomes.

**Margolis et al.: Early Career Experiences and Later Career Outcomes: Comparing the United States, France and Germany** **MargolisEtAl2001**

---

David N. Margolis, Véronique Simonnet, and Lars Vilhuber. “Early Career Experiences and Later Career Outcomes: Comparing the United States, France and Germany”. In: *Vierteljahrshefte zur Wirtschaftsforschung* 70.1 (2001), pp. 31–38. DOI: [10.3790/vjh.70.1.31](https://doi.org/10.3790/vjh.70.1.31).

Abstract: This paper explores the links between individuals’ early career experiences and their labor market outcomes 5 to 20 years later using data

from France, (western) Germany, and the United States. Relative to most of the literature, we consider a large set of measures of men’s early career experiences and later career outcomes. Our results differ significantly across countries. Labor market outcomes in Germany are consistent with a dual labor market model. In the case of American workers, either the market learns about unobservable worker characteristics over time or the implicit contracts established at the start of the career are increasingly renegotiated over time. Unobserved heterogeneity in individuals’ networks of labor market contacts is consistent with our results for France. These results reflect optimal firm responses to the different institutional environments in each country in the presence of ex ante imperfect information concerning young workers.

**Vilhuber: La spécificité de la formation en milieu de travail : un survol des contributions théoriques et empiriques récentes,**

**Vilhuber2001**

---

Lars Vilhuber. “La spécificité de la formation en milieu de travail : un survol des contributions théoriques et empiriques récentes,” in: *L’Actualité économique, Revue d’analyse économique* 77.1 (2001).

**Vilhuber: Continuous Training and sectoral mobility in Germany: Evidence from the 90s**

**Vilhuber99a**

---

Lars Vilhuber. “Continuous Training and sectoral mobility in Germany: Evidence from the 90s”. In: *Vierteljahresheft für Wirtschaftsforschung* 68.2 (1999), pp. 209–214. URL: <http://hdl.handle.net/10419/141240>.

Abstract: see Vilhuber99b.

## 2 Proceedings

**John M. Abowd et al.: Disclosure Limitation and Confidentiality Protection in Linked Data**

**abowd\_disclosure\_2021**

---

John M. Abowd, Ian M. Schmutte, and Lars Vilhuber. “Disclosure Limitation and Confidentiality Protection in Linked Data”. In: *Administrative Records for Survey Methodology*. Ed. by Asaph Young Chun and Michael D. Larsen. Survey Research Methods & Sampling. Wiley, 2021. ISBN: 978-1-119-27204-

5. URL: <https://www.wiley.com/en-ca/Administrative+Records+for+Survey+Methodology-p-9781119272045>.

Abstract: Confidentiality protection for linked administrative data is a combination of access modalities and statistical disclosure limitation. We review traditional statistical disclosure limitation methods and newer methods based on synthetic data, input noise infusion and formal privacy. We discuss how these methods are integrated with access modalities by providing three detailed examples. The first example is the linkages in the Health and Retirement Study to Social Security Administration data. The second example is the linkage of the Survey of Income and Program Participation to administrative data from the Internal Revenue Service and the Social Security Administration. The third example is the Longitudinal Employer-Household Dynamics data, which links state unemployment insurance records for workers and firms to a wide variety of censuses and surveys at the U.S. Census Bureau. For examples, we discuss access modalities, disclosure limitation methods, the effectiveness of those methods, and the resulting analytical validity. The final sections discuss recent advances in access modalities for linked administrative data.

### **Villhuber: Making Confidential Data Part of Reproducible Research NAP25305**

---

Lars Villhuber. “Making Confidential Data Part of Reproducible Research”. In: *Methods to Foster Transparency and Reproducibility of Federal Statistics: Proceedings of a Workshop*. Ed. by National Academies of Sciences, Engineering, and Medicine. National Academies of Sciences, Engineering, and Medicine; Michael Cohen (Rapporteur). Washington, DC: The National Academies Press, 2019, pp. 63–66. ISBN: 978-0-309-48629-3. DOI: [10.17226/25305](https://doi.org/10.17226/25305).

Abstract: In 2014 the National Science Foundation (NSF) provided support to the National Academies of Sciences, Engineering, and Medicine for a series of Forums on Open Science in response to a government-wide directive to support increased public access to the results of research funded by the federal government. However, the breadth of the work resulting from the series precluded a focus on any specific topic or discussion about how to improve public access. Thus, the main goal of the Workshop on Transparency and Reproducibility in Federal Statistics was to develop some understanding of

what principles and practices are, or would be, supportive of making federal statistics more understandable and reviewable, both by agency staff and the public. This publication summarizes the presentations and discussions from the workshop.

**Pistner et al.: Synthetic Data via Quantile Regression for Heavy-Tailed and Heteroskedastic Data**

**PistnerSlavkovicVilhuber:PSD:2018**

---

Michelle Pistner, Aleksandra Slavković, and Lars Vilhuber. “Synthetic Data via Quantile Regression for Heavy-Tailed and Heteroskedastic Data”. In: *Privacy in Statistical Databases*. Ed. by Josep Domingo-Ferrer and Francisco Montes. 2018. DOI: [10.1007/978-3-319-99771-1\\_7](https://doi.org/10.1007/978-3-319-99771-1_7).

**Haney et al.: Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics**

**HaneySIGMOD2017**

---

Samuel Haney, Ashwin Machanavajjhala, John M. Abowd, Matthew Graham, Mark Kutzbach, and Lars Vilhuber. “Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics”. In: *Proceedings of the 2017 International Conference on Management of Data*. SIGMOD ’17. ACM, 2017, pp. 1339–1354. DOI: [10.1145/3035918.3035940](https://doi.org/10.1145/3035918.3035940).

Abstract: National statistical agencies around the world publish tabular summaries based on combined employer-employee (ER-EE) data. The privacy of both individuals and business establishments that feature in these data are protected by law in most countries. These data are currently released using a variety of statistical disclosure limitation (SDL) techniques that do not reveal the exact characteristics of particular employers and employees, but lack provable privacy guarantees limiting inferential disclosures. In this work, we present novel algorithms for releasing tabular summaries of linked ER-EE data with formal, provable guarantees of privacy. We show that state-of-the-art differentially private algorithms add too much noise for the output to be useful. Instead, we identify the privacy requirements mandated by current interpretations of the relevant laws, and formalize them using the Pufferfish framework. We then develop new privacy definitions that are customized to ER-EE data and satisfy the statutory privacy requirements. We implement the experiments in this paper on production data gathered by the U.S. Census Bureau. An empirical evaluation of utility for these data shows that for



reasonable values of the privacy-loss parameter  $\epsilon \geq 1$ , the additive error introduced by our provably private algorithms is comparable, and in some cases better, than the error introduced by existing SDL techniques that have no provable privacy guarantees. For some complex queries currently published, however, our algorithms do not have utility comparable to the existing traditional SDL algorithms. Those queries are fodder for future research.

**Jörg Drechsler et al.: Synthetic Longitudinal Business Databases  
for International Comparisons** **psd2014b**

---

Jörg Drechsler and Lars Vilhuber. “Synthetic Longitudinal Business Databases for International Comparisons”. In: *Privacy in Statistical Databases*. Ed. by Josep Domingo-Ferrer. Vol. 8744. Lecture Notes in Computer Science. Springer International Publishing, 2014, pp. 243–252. ISBN: 978-3-319-11256-5. DOI: [10.1007/978-3-319-11257-2\\_19](https://doi.org/10.1007/978-3-319-11257-2_19).

Abstract: International comparison studies on economic activity are often hampered by the fact that access to business microdata is very limited on an international level. A recently launched project tries to overcome these limitations by improving access to Business Censuses from multiple countries based on synthetic data. Starting from the synthetic version of the longitudinally edited version of the U.S. Business Register (the Longitudinal Business Database, LBD), the idea is to create similar data products in other countries by applying the synthesis methodology developed for the LBD to generate synthetic replicates that could be distributed without confidentiality concerns. In this paper we present some first results of this project based on German business data collected at the Institute for Employment Research.

**Lagoze et al.: CED2AR: The Comprehensive Extensible Data Documentation and Access Repository** **LagozeJCDL2014**

---

Carl Lagoze, Lars Vilhuber, Jeremy Williams, Benjamin Perry, and William C. Block. “CED2AR: The Comprehensive Extensible Data Documentation and Access Repository”. In: *ACM/IEEE Joint Conference on Digital Libraries (JCDL 2014)*. Presented at the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2014). ACM/IEEE. London, United Kingdom, 2014.

Abstract: Social science researchers increasingly make use of data that is confidential because it contains linkages to the identities of people, corpo-

rations, etc. The value of this data lies in the ability to join the identifiable entities with external data such as genome data, geospatial information, and the like. However, the confidentiality of this data is a barrier to its utility and curation, making it difficult to fulfill US federal data management mandates and interfering with basic scholarly practices such as validation and reuse of existing results. We describe the complexity of the relationships among data that span a public and private divide. We then describe our work on the CED2AR prototype, a first step in providing researchers with a tool that spans this divide and makes it possible for them to search, access, and cite that data.

**Miranda et al.: Using Partially Synthetic Data to Replace Suppression in the Business Dynamics Statistics: Early Results**    **psd2014a**

---

Javier Miranda and Lars Vilhuber. “Using Partially Synthetic Data to Replace Suppression in the Business Dynamics Statistics: Early Results”. In: *Privacy in Statistical Databases*. Ed. by Josep Domingo-Ferrer. Vol. 8744. Lecture Notes in Computer Science. Springer International Publishing, 2014, pp. 232–242. ISBN: 978-3-319-11256-5. DOI: [10.1007/978-3-319-11257-2\\_18](https://doi.org/10.1007/978-3-319-11257-2_18).

Abstract: The Business Dynamics Statistics is a product of the U.S. Census Bureau that provides measures of business openings and closings, and job creation and destruction, by a variety of cross-classifications (firm and establishment age and size, industrial sector, and geography). Sensitive data are currently protected through suppression. However, as additional tabulations are being developed, at ever more detailed geographic levels, the number of suppressions increases dramatically. This paper explores the option of providing public-use data that are analytically valid and without suppressions, by leveraging synthetic data to replace observations in sensitive cells.

**Lagoze et al.: Encoding Provenance of Social Science Data: Integrating PROV with DDI**    **LagozeEtAl2013**

---

Carl Lagoze, William C. Block, Jeremy Williams, and Lars Vilhuber. “Encoding Provenance of Social Science Data: Integrating PROV with DDI”. In: *5th Annual European DDI User Conference*. 2013.

Abstract: Provenance is a key component of evaluating the integrity and reusability of data for scholarship. While recording and providing access provenance has always been important, it is even more critical in the web environment in which data from distributed sources and of varying integrity can be combined and derived. The PROV model, developed under the auspices of the W3C, is a foundation for semantically-rich, interoperable, and web-compatible provenance metadata. We report on the results of our experimentation with integrating the PROV model into the DDI metadata for a complex, but characteristic, example social science data. We also present some preliminary thinking on how to visualize those graphs in the user interface.

**Lagoze et al.: Encoding Provenance Metadata for Social Science Datasets**  
**LagozeEtAl2013b**

---

Carl Lagoze, Jeremy Williams, and Lars Vilhuber. “Encoding Provenance Metadata for Social Science Datasets”. In: *Metadata and Semantics Research*. Ed. by Emmanouel Garoufallou and Jane Greenberg. Vol. 390. Communications in Computer and Information Science. Springer International Publishing, 2013, pp. 123–134. ISBN: 978-3-319-03436-2. DOI: [10.1007/978-3-319-03437-9\\_13](https://doi.org/10.1007/978-3-319-03437-9_13).

Abstract: Recording provenance is a key requirement for data-centric scholarship, allowing researchers to evaluate the integrity of source data sets and reproduce, and thereby, validate results. Provenance has become even more critical in the web environment in which data from distributed sources and of varying integrity can be combined and derived. Recent work by the W3C on the PROV model provides the foundation for semantically-rich, interoperable, and web-compatible provenance metadata. We apply that model to complex, but characteristic, provenance examples of social science data, describe scenarios that make scholarly use of those provenance descriptions, and propose a manner for encoding this provenance metadata within the widely-used DDI metadata standard.

**John M. Abowd et al.: A Proposed Solution to the Archiving and Curation of Confidential Scientific Inputs**  
**AbowdVilhuberBlock2012**

---

John M. Abowd, Lars Vilhuber, and William Block. “A Proposed Solution to the Archiving and Curation of Confidential Scientific Inputs”. In: *Privacy*

in *Statistical Databases*. Ed. by Josep Domingo-Ferrer and Ilenia Tinnirello. Vol. 7556. Lecture Notes in Computer Science. Springer, 2012, pp. 216–225. ISBN: 978-3-642-33626-3. DOI: [10.1007/978-3-642-33627-0\\_17](https://doi.org/10.1007/978-3-642-33627-0_17).

**Dostie et al.: Using linked employer-employee data to investigate the speed of adjustment in downsizing firms in Canada and the US**  
**DostieMcKinneyVilhuber2009**

---

Benoit Dostie, Kevin L. McKinney, and Lars Vilhuber. “Using linked employer-employee data to investigate the speed of adjustment in downsizing firms in Canada and the US”. In: *International Census Research Data Center Conference*. Ithaca, NY, 2009.

Abstract: When firms are faced with a demand shock, adjustment can take many forms. Firms can adjust physical capital, human capital, or both. The speed of adjustment may differ as well: costs of adjustment, the type of shock, the legal and economic environment all matter. In this paper, we focus on firms that downsized between 1992 and 1997, but ultimately survive, and investigate how the human capital distribution within a firm influences the speed of adjustment, *ceteris paribus*. In other words, when do firms use mass layoffs instead of attrition to adjust the level of employment. We combine worker-level wage records and measures of human capital with firm-level characteristics of the production function, and use levels and changes in these variables to characterize the choice of adjustment method and speed. Firms are described/compared up to 9 years prior to death. We also consider how workers fare after leaving downsizing firms, and analyze if observed differences in post-separation outcomes of workers provide clues to the choice of adjustment speed.

File: [McKinneyVilhuber2006-ESEM2006.pdf](#) : [/home/vilhuber/Textes/Papers/Census/CAFE-displacement-death/releases/2006-05-24/McKinneyVilhuber2006-ESEM2006.pdf](#):PDF.

**J. Abowd et al.: How Protective Are Synthetic Data**  
**AbowdVilhuber2008**

---

John Abowd and Lars Vilhuber. “How Protective Are Synthetic Data”. In: *Privacy in Statistical Database*. Ed. by Josep Domingo-Ferrer and Yücel

Saygin. Vol. 5262. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, pp. 239–246. DOI: [10.1007/978-3-540-87471-3\\_20](https://doi.org/10.1007/978-3-540-87471-3_20).

**Dragoset et al.: How Did Universal Primary Education Affect Returns to Education and Labor Market Participation in Uganda?**  
**DragosetVilhuber2008**

---

Lisa Dragoset and Lars Vilhuber. “How Did Universal Primary Education Affect Returns to Education and Labor Market Participation in Uganda?” In: *YOUTH IN AFRICA’S LABOR MARKET*. Ed. by M Garcia and J Fares. 1818 H ST NW, WASHINGTON, DC 20433 USA: WORLD BANK INST, 2008, pp. 263–280. ISBN: 978-0-8213-6885-5. DOI: [10.1596/978-0-8213-6884-8\\_ch11](https://doi.org/10.1596/978-0-8213-6884-8_ch11).

**K. L. McKinney et al.: Using linked employer-employee data to investigate the speed of adjustment in downsizing firms**  
**McKinneyVilhuber2006**

---

Kevin L. McKinney and Lars Vilhuber. “Using linked employer-employee data to investigate the speed of adjustment in downsizing firms”. In: *Conference on the Analysis of Firms and Employees (CAFE)*. Nuremberg, Germany, 2006.

Abstract: When firms are faced with a demand shock, adjustment can take many forms. Firms can adjust physical capital, human capital, or both. The speed of adjustment may differ as well: costs of adjustment, the type of shock, the legal and economic environment all matter. In this paper, we focus on firms that downsized between 1992 and 1997, but ultimately survive, and investigate how the human capital distribution within a firm influences the speed of adjustment, *ceteris paribus*. In other words, when do firms use mass layoffs instead of attrition to adjust the level of employment. We combine worker-level wage records and measures of human capital with firm-level characteristics of the production function, and use levels and changes in these variables to characterize the choice of adjustment method and speed. Firms are described/compared up to 9 years prior to death. We also consider how workers fare after leaving downsizing firms, and analyze if observed differences in post-separation outcomes of workers provide clues to the choice of adjustment speed.

File: McKinneyVilhuber2006-ESEM2006.pdf : /home/vilhuber/Textes/Papers/Census/CAFE-displacement-death/releases/2006-05-24/McKinneyVilhuber2006-ESEM2006.pdf:PDF.

### 3 Book chapters

**I. M. Schmutte et al.: Balancing Privacy and Data Usability: An Overview of Disclosure Avoidance Methods**

**SchmutteVilhuber2020**

---

Ian M. Schmutte and Lars Vilhuber. “Balancing Privacy and Data Usability: An Overview of Disclosure Avoidance Methods”. In: *Handbook on Using Administrative Data for Research and Evidence-Based Policy*. Ed. by Shawn Cole, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber. (pre-print version). Abdul Latif Jameel Poverty Action Lab, 2020. ISBN: TBD. URL: <https://admindatahandbook.mit.edu/book/latest/discavoid.html> (visited on 10/18/2020).

File: Snapshot:/Users/larsvilhuber/Zotero/storage/V828RV3I/admindatahandbook.mit.edu.html:text/html.

**Shen et al.: Physically Protecting Sensitive Data**

**ShenVilhuber2020**

---

Jim H. Shen and Lars Vilhuber. “Physically Protecting Sensitive Data”. In: *Handbook on Using Administrative Data for Research and Evidence-Based Policy*. Ed. by Shawn Cole, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber. (pre-print version). Abdul Latif Jameel Poverty Action Lab, 2020. ISBN: TBD. URL: <https://admindatahandbook.mit.edu/book/latest/security.html> (visited on 10/18/2020).

File: Snapshot:/Users/larsvilhuber/Zotero/storage/V828RV3I/admindatahandbook.mit.edu.html:text/html.

**John M. Abowd et al.: The link between human capital, mass layoffs, and firm deaths**

**AbowdEtAl2009c**

---

John M. Abowd, Kevin L. McKinney, and Lars Vilhuber. “The link between human capital, mass layoffs, and firm deaths”. In: *Producer Dynamics: New*

*Evidence from Micro Data*. Ed. by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. University of Chicago Press, 2009. ISBN: 978-0-226-17256-9. URL: <http://www.nber.org/chapters/c0497/>.

File: [AbowdMcKinneyVilhuber2005.pdf:L/LEHD/AbowdMcKinneyVilhuber2005.pdf:PDF](#).

---

**John M. Abowd et al.: The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators    AbowdEtAl2009**

---

John M. Abowd, Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon D. Woodcock. “The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators”. In: *Producer Dynamics: New Evidence from Micro Data*. Ed. by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts. University of Chicago Press, 2009. ISBN: 978-0-226-17256-9. URL: <http://www.nber.org/chapters/c0485>.

File: [AbowdStephensVilhuber2005-LEHD-final.pdf:L/LEHD/AbowdStephensVilhuber2005-LEHD-final.pdf:PDF](#).

---

**Vilhuber: Adjusting Imperfect Data: Overview and Case Studies    NBERc2366**

---

Lars Vilhuber. “Adjusting Imperfect Data: Overview and Case Studies”. In: *The Structure of Wages: An International Comparison*. Ed. by Edward P. Lazear and Kathryn L. Shaw. University of Chicago Press, 2009, pp. 59–80. URL: <http://www.nber.org/chapters/c2366>.

---

**Margolis et al.: Early Career Experiences and Later Career Outcomes: An International Comparison    MargolisEtAl2004**

---

David N. Margolis, Erik Plug, Véronique Simonnet, and Lars Vilhuber. “Early Career Experiences and Later Career Outcomes: An International Comparison”. In: *Human Capital Over The Life Cycle - A European Perspective*. Ed. by Catherine Sofer. London: Edward Elgar, 2004. Chap. 5, pp. 90–117.

## 4 Working papers and other documents

**Alam et al.: Applying Data Synthesis for Longitudinal Business Data across Three Countries** **alam2020applying**

---

M. Jahangir Alam, Benoit Dostie, Jörg Drechsler, and Lars Vilhuber. *Applying Data Synthesis for Longitudinal Business Data across Three Countries*. arXiv:2008.02246. 2020. URL: <https://arxiv.org/abs/2008.02246>.

**Lange et al.: Consumer expectations around COVID-19: Evolution over time** **langevilhuber202005**

---

Fabian Lange and Lars Vilhuber. *Consumer expectations around COVID-19: Evolution over time*. Online. Labor Dynamics Institute, 2020. URL: [https://labordynamicsinstitute.github.io/gcs\\_covid19\\_expectations/text/analysis\\_week5/](https://labordynamicsinstitute.github.io/gcs_covid19_expectations/text/analysis_week5/).

**J. Abowd et al.: Introductory Readings in Formal Privacy for Economists** **formalprivacy:zenodo:2019**

---

John Abowd, Ian Schmutte, William Sexton, and Lars Vilhuber. *Introductory Readings in Formal Privacy for Economists*. Tech. rep. Supported by Alfred P. Sloan Foundation Grant G-2015-13903 and NSF Grant SES-1131848. 2019. DOI: [10.5281/zenodo.2621345](https://doi.org/10.5281/zenodo.2621345).

**Lagoze et al.: metajelo: A metadata package for journals to support external linked objects** **VilhuberLagoze2019**

---

Carl Lagoze and Lars Vilhuber. *metajelo: A metadata package for journals to support external linked objects*. Document 52. Labor Dynamics Institute, 2019. URL: <http://digitalcommons.ilr.cornell.edu/ldi/52/>.

Abstract: We propose a metadata package that is intended to provide academic journals with a lightweight means of registering, at the time of publication, the existence and disposition of supplementary materials. Information about the supplementary materials is, in most cases, critical for the reproducibility and replicability of scholarly results. In many instances, these materials are curated by a third party, which may or may not follow developing standards for the identification and description of those materials. As such, the vocabulary described here complements existing initiatives that



specify vocabularies to describe the supplementary materials or the repositories and archives in which they have been deposited. Where possible, it reuses elements of relevant other vocabularies, facilitating coexistence with them. Furthermore, it provides an “at publication” record of reproducibility characteristics of a particular article that has been selected for publication. The proposed metadata package documents the key characteristics that journals care about in the case of supplementary materials that are held by third parties: existence, accessibility, and permanence. It does so in a robust, time-invariant fashion at the time of publication, when the editorial decisions are made. It also allows for better documentation of less accessible (non-public data), by treating it symmetrically from the point of view of the journal, therefore increasing the transparency of what up until now has been very opaque.

**Vilhuber et al.: Outcomes Report of the Cornell Node of the NSF-Census Research Network** **VilhuberBlock2019**

---

Lars Vilhuber and William Block. *Outcomes Report of the Cornell Node of the NSF-Census Research Network*. Report. 2019. URL: <https://hdl.handle.net/1813/65011> (visited on 04/10/2019).

Abstract: Description and List of Outcomes of the Cornell node of the NSF-Census Research Network.

File: [FullTextPDF:/home/vilhuber/Zotero/storage/AGTYI8RG/VilhuberandBlock-2019-OutcomesreportCornellNodeoftheNSF-CensusR.pdf:application/pdf;Snapshot:/home/vilhuber/Zotero/storage/JPN6HF5W/65011.html:text/html](#).

**John M. Abowd et al.: Disclosure Limitation and Confidentiality Protection in Linked Data** **RePEc:cen:wpaper:18-07**

---

John M. Abowd, Ian M. Schmutte, and Lars Vilhuber. *Disclosure Limitation and Confidentiality Protection in Linked Data*. Working Papers 18-07. Center for Economic Studies, U.S. Census Bureau, 2018. URL: <https://ideas.repec.org/p/cen/wpaper/18-07.html>.

Abstract: Confidentiality protection for linked administrative data is a combination of access modalities and statistical disclosure limitation. We review traditional statistical disclosure limitation methods and newer methods

based on synthetic data, input noise infusion and formal privacy. We discuss how these methods are integrated with access modalities by providing three detailed examples. The first example is the linkages in the Health and Retirement Study to Social Security Administration data. The second example is the linkage of the Survey of Income and Program Participation to administrative data from the Internal Revenue Service and the Social Security Administration. The third example is the Longitudinal Employer-Household Dynamics data, which links state unemployment insurance records for workers and firms to a wide variety of censuses and surveys at the U.S. Census Bureau. For examples, we discuss access modalities, disclosure limitation methods, the effectiveness of those methods, and the resulting analytical validity. The final sections discuss recent advances in access modalities for linked administrative data.

#### **King et al.: Data Sharing Governance and Management**

**adrf-working**

---

Monica King and Working Group Participants. *Data Sharing Governance and Management*. Working Group Report. Administrative Data Research Facilities Network, 2018. URL: [https://docs.wixstatic.com/ugd/e35e20\\_33eaf93b681846248aac487e6f23e85c.pdf](https://docs.wixstatic.com/ugd/e35e20_33eaf93b681846248aac487e6f23e85c.pdf).

#### **Vilhuber: LEHD Infrastructure S2014 files in the FSRDC**

**RePEc:cen:wpaper:18-27**

---

Lars Vilhuber. *LEHD Infrastructure S2014 files in the FSRDC*. Working Papers 18-27. Center for Economic Studies, U.S. Census Bureau, 2018. URL: <https://ideas.repec.org/p/cen/wpaper/18-27.html>.

Abstract: The Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau, with the support of several national research agencies, maintains a set of infrastructure files using administrative data provided by state agencies, enhanced with information from other administrative data sources, demographic and economic (business) surveys and censuses. The LEHD Infrastructure Files provide a detailed and comprehensive picture of workers, employers, and their interaction in the U.S. economy. This document describes the structure and content of the 2014 Snapshot of the LEHD Infrastructure files as they are made available in the Census Bureau's secure and restricted-access Research Data Center network. The

document attempts to provide a comprehensive description of all researcher-accessible files, of their creation, and of any modifications made to the files to facilitate researcher access.

**Vilhuber: Reproducibility and Replicability in Economics**  
**vilhuber2018nap**

---

Lars Vilhuber. *Reproducibility and Replicability in Economics*. Commissioned Paper. National Academies of Sciences, Engineering, and Medicine, 2018. URL: <https://www.nap.edu/resource/25303/Reproducibility%20in%20Economics.pdf>.

**Cloutier et al.: Understanding the effect of procedural justice on psychological distress**  
**CloutierVilhuberLDI2017**

---

Julie Cloutier, Lars Vilhuber, Denis Harrisson, and Vanessa Béland-Ouellette. *Understanding the effect of procedural justice on psychological distress*. Document 35. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/35/>.

Abstract: Studies on the effect of procedural justice on psychological distress present conflicting results. Drawing on instrumental and relational perspectives of justice, we test the hypothesis that the perception of procedural justice influences the level of workers' psychological distress. Using a number of validated instruments to collected data from 659 workers in three call centers, we use OLS regressions and Hayes' PROCESS tool to show that the perception of procedural justice has a direct, unique, and independent effect on psychological distress. The perception of procedural justice has no instrumental role, the key mechanism being the relational role, suggesting that perceived injustice influences psychological distress because it threatens self-esteem. Distributive justice perceptions (recognition, promotions, job security) are not associated with psychological distress, calling into question Siegrist's model. Our findings suggest that perceived procedural justice provides workers better evidence of the extent to which they are valued and appreciated members of their organizations than do perceptions of distributive justice. The results highlight the greater need for workers to be valued and appreciated for who they are (consideration and esteem), rather than for what they do for their organization (distributive justice of rewards).

**Foote et al.: Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings**

**RePEc:cen:wpaper:17-49**

---

Andrew Foote, Mark J. Kutzbach, and Lars Vilhuber. *Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings*. Working Papers 17-49. Center for Economic Studies, U.S. Census Bureau, 2017. URL: <https://ideas.repec.org/p/cen/wpaper/17-49.html>.

Abstract: This paper evaluates the use of commuting zones as a local labor market definition. We revisit Tolbert and Sizer (1996) and demonstrate the sensitivity of definitions to two features of the methodology. We show how these features impact empirical estimates using a well-known application of commuting zones. We conclude with advice to researchers using commuting zones on how to demonstrate the robustness of empirical findings to uncertainty in definitions.

**Foote et al.: Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings** **handle:1813:52649**

---

Andrew Foote, Mark J. Kutzbach, and Lars Vilhuber. *Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings*. Preprint 1813:52649. NSF Census Research Network - NCRN-Cornell, 2017. URL: <http://hdl.handle.net/1813/52649>.

Abstract: This paper evaluates the use of commuting zones as a local labor market definition. We revisit Tolbert and Sizer (1996) and demonstrate the sensitivity of definitions to two features of the methodology. We show how these features impact empirical estimates using a well-known application of commuting zones. We conclude with advice to researchers using commuting zones on how to demonstrate the robustness of empirical findings to uncertainty in definitions. The analysis, conclusions, and opinions expressed herein are those of the author(s) alone and do not necessarily represent the views of the U.S. Census Bureau or the Federal Deposit Insurance Corporation. All results have been reviewed to ensure that no confidential information is disclosed, and no confidential data was used in this paper. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Much of the work developing this paper occurred while Mark Kutzbach was an employee of the U.S. Census Bureau.

**Foote et al.: Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings**  
**ldi:2017:45**

---

Andrew Foote, Mark J. Kutzbach, and Lars Vilhuber. *Recalculating - How Uncertainty in Local Labor Market Definitions Affects Empirical Findings*. Document 45. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/45/>.

Abstract: This paper evaluates the use of commuting zones as a local labor market definition. We revisit Tolbert and Sizer (1996) and demonstrate the sensitivity of definitions to two features of the methodology. We show how these features impact empirical estimates using a well-known application of commuting zones. We conclude with advice to researchers using commuting zones on how to demonstrate the robustness of empirical findings to uncertainty in definitions. The analysis, conclusions, and opinions expressed herein are those of the author(s) alone and do not necessarily represent the views of the U.S. Census Bureau or the Federal Deposit Insurance Corporation. All results have been reviewed to ensure that no confidential information is disclosed, and no confidential data was used in this paper. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Much of the work developing this paper occurred while Mark Kutzbach was an employee of the U.S. Census Bureau.

**Green et al.: Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files**  
**RePEc:cen:wpaper:17-34**

---

Andrew S. Green, Mark J. Kutzbach, and Lars Vilhuber. *Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files*. Working Papers 17-34. Center for Economic Studies, U.S. Census Bureau, 2017. URL: <https://ideas.repec.org/p/cen/wpaper/17-34.html>.

Abstract: Commuting flows and workplace employment data have a wide constituency of users including urban and regional planners, social science and transportation researchers, and businesses. The U.S. Census Bureau releases two, national data products that give the magnitude and characteristics of home to work flows. The American Community Survey (ACS) tabulates households' responses on employment, workplace, and commuting be-

havior. The Longitudinal Employer-Household Dynamics (LEHD) program tabulates administrative records on jobs in the LEHD Origin-Destination Employment Statistics (LODES). Design differences across the datasets lead to divergence in a comparable statistic: county-to-county aggregate commute flows. To understand differences in the public use data, this study compares ACS and LEHD source files, using identifying information and probabilistic matching to join person and job records. In our assessment, we compare commuting statistics for job frames linked on person, employment status, employer, and workplace and we identify person and job characteristics as well as design features of the data frames that explain aggregate differences. We find a lower rate of within-county commuting and farther commutes in LODES. We attribute these greater distances to differences in workplace reporting and to uncertainty of establishment assignments in LEHD for workers at multi-unit employers. Minor contributing factors include differences in residence location and ACS workplace edits. The results of this analysis and the data infrastructure developed will support further work to understand and enhance commuting statistics in both datasets.

**Green et al.: Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files**  
**handle:1813:52611**

---

Andrew S. Green, Mark J. Kutzbach, and Lars Vilhuber. *Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files*. Preprint 1813:52611. NSF Census Research Network - NCRN-Cornell, 2017. URL: <http://hdl.handle.net/1813/52611>.

Abstract: Commuting flows and workplace employment data have a wide constituency of users including urban and regional planners, social science and transportation researchers, and businesses. The U.S. Census Bureau releases two, national data products that give the magnitude and characteristics of home to work flows. The American Community Survey (ACS) tabulates households' responses on employment, workplace, and commuting behavior. The Longitudinal Employer-Household Dynamics (LEHD) program tabulates administrative records on jobs in the LEHD Origin-Destination Employment Statistics (LODES). Design differences across the datasets lead to divergence in a comparable statistic: county-to-county aggregate commute

flows. To understand differences in the public use data, this study compares ACS and LEHD source files, using identifying information and probabilistic matching to join person and job records. In our assessment, we compare commuting statistics for job frames linked on person, employment status, employer, and workplace and we identify person and job characteristics as well as design features of the data frames that explain aggregate differences. We find a lower rate of within-county commuting and farther commutes in LODES. We attribute these greater distances to differences in workplace reporting and to uncertainty of establishment assignments in LEHD for workers at multi-unit employers. Minor contributing factors include differences in residence location and ACS workplace edits. The results of this analysis and the data infrastructure developed will support further work to understand and enhance commuting statistics in both datasets.

**Green et al.: Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files** **ldi:2017:38**

---

Andrew S. Green, Mark J. Kutzbach, and Lars Vilhuber. *Two Perspectives on Commuting: A Comparison of Home to Work Flows Across Job-Linked Survey and Administrative Files*. Document 38. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/38/>.

Abstract: Commuting flows and workplace employment data have a wide constituency of users including urban and regional planners, social science and transportation researchers, and businesses. The U.S. Census Bureau releases two, national data products that give the magnitude and characteristics of home to work flows. The American Community Survey (ACS) tabulates households' responses on employment, workplace, and commuting behavior. The Longitudinal Employer-Household Dynamics (LEHD) program tabulates administrative records on jobs in the LEHD Origin-Destination Employment Statistics (LODES). Design differences across the datasets lead to divergence in a comparable statistic: county-to-county aggregate commute flows. To understand differences in the public use data, this study compares ACS and LEHD source files, using identifying information and probabilistic matching to join person and job records. In our assessment, we compare commuting statistics for job frames linked on person, employment status,



employer, and workplace and we identify person and job characteristics as well as design features of the data frames that explain aggregate differences. We find a lower rate of within-county commuting and farther commutes in LODES. We attribute these greater distances to differences in workplace reporting and to uncertainty of establishment assignments in LEHD for workers at multi-unit employers. Minor contributing factors include differences in residence location and ACS workplace edits. The results of this analysis and the data infrastructure developed will support further work to understand and enhance commuting statistics in both datasets.

**Haney et al.: Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics** **handle:1813:49652**

---

Samuel Haney, Ashwin Machanavajjhala, John M Abowd, Matthew Graham, and Mark Kutzbach. *Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics*. Preprint 1813:49652. Cornell University, 2017. URL: <http://hdl.handle.net/1813/49652>.

Abstract: Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics Haney, Samuel; Machanavajjhala, Ashwin; Abowd, John M; Graham, Matthew; Kutzbach, Mark National statistical agencies around the world publish tabular summaries based on combined employeremployee (ER-EE) data. The privacy of both individuals and business establishments that feature in these data are protected by law in most countries. These data are currently released using a variety of statistical disclosure limitation (SDL) techniques that do not reveal the exact characteristics of particular employers and employees, but lack provable privacy guarantees limiting inferential disclosures. In this work, we present novel algorithms for releasing tabular summaries of linked ER-EE data with formal, provable guarantees of privacy. We show that state-of-the-art differentially private algorithms add too much noise for the output to be useful. Instead, we identify the privacy requirements mandated by current interpretations of the relevant laws, and formalize them using the Pufferfish framework. We then develop new privacy definitions that are customized to ER-EE data and satisfy the statutory privacy requirements. We implement the experiments in this paper on production data gathered by the U.S. Census Bureau. An empirical evaluation of utility for these data shows that for reasonable values of the privacy-loss parameter  $\epsilon \geq 1$ , the additive error introduced by our provably private algo-



rithms is comparable, and in some cases better, than the error introduced by existing SDL techniques that have no provable privacy guarantees. For some complex queries currently published, however, our algorithms do not have utility comparable to the existing traditional &quot;This Article is brought to you for free and open access by the Centers, Institutes, Programs at DigitalCommons@ILR. It has been accepted for inclusion in Labor Dynamics Institute by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.&quot;

**K. L. McKinney et al.: Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On The Map**  
**RePEc:cen:wpaper:17-71**

---

Kevin L. McKinney, Andrew S. Green, Lars Vilhuber, and John M. Abowd. *Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On The Map*. Working Papers 17-71. Center for Economic Studies, U.S. Census Bureau, 2017. URL: <https://ideas.repec.org/p/cen/wpaper/17-71.html>.

**Abstract:** We report results from the first comprehensive total quality evaluation of five major indicators in the U.S. Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) Program Quarterly Workforce Indicators (QWI): total employment, beginning-of-quarter employment, full-quarter employment, total payroll, and average monthly earnings of full-quarter employees. Beginning-of-quarter employment is also the main tabulation variable in the LEHD Origin-Destination Employment Statistics (LODES) workplace reports as displayed in OnTheMap (OTM). The evaluation is conducted by generating multiple threads of the edit and imputation models used in the LEHD Infrastructure File System. These threads conform to the Rubin (1987) multiple imputation model, with each thread or imputation being the output of formal probability models that address coverage, edit, and imputation errors. Design-based sampling variability and finite population corrections are also included in the evaluation. We derive special formulas for the Rubin total variability and its components that are consistent with the disclosure avoidance system used for QWI and LODES/OTM workplace reports. These formulas allow us to publish the complete set of detailed total

quality measures for QWI and LODES. The analysis reveals that the five publication variables under study are estimated very accurately for tabulations involving at least 10 jobs. Tabulations involving three to nine jobs have quality in the range generally deemed acceptable. Tabulations involving zero, one or two jobs, which are generally suppressed in the QWI and synthesized in LODES, have substantial total variability but their publication in LODES allows the formation of larger custom aggregations, which will in general have the accuracy estimated for tabulations in the QWI based on a similar number of workers.

---

**Vilhuber et al.: Proceedings from the Synthetic LBD International Seminar** **ProceedingsSynLBD2017**

---

Lars Vilhuber, Saki Kinney, and Ian Schmutte. *Proceedings from the Synthetic LBD International Seminar*. Document 44. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/44/>.

Abstract: On May 9, 2017, we hosted a seminar to discuss the conditions necessary to implement the SynLBD approach with interested parties, with the goal of providing a straightforward toolkit to implement the same procedure on other data. The proceedings summarize the discussions during the workshop. Funding for the workshop was provided by the National Science Foundation (Grants 1012593; 1131848) and the Alfred P. Sloan Foundation (G-2015-13903). Organizational support was provided by the Labor Dynamics Institute at Cornell University.

---

**Vilhuber et al.: Making Confidential Data Part of Reproducible Research** **VilhuberLagozeLDI2017**

---

Lars Vilhuber and Carl Lagoze. *Making Confidential Data Part of Reproducible Research*. Document 41. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/41/>.

---

**Vilhuber et al.: Proceedings from the 2016 NSF-Sloan Workshop on Practical Privacy** **handle:1813:46197**

---

Lars Vilhuber and Ian Schmutte. *Proceedings from the 2016 NSF-Sloan Work-*

*shop on Practical Privacy*. Preprint 1813:46197. Cornell University, 2017. URL: <http://hdl.handle.net/1813/46197>.

Abstract: Proceedings from the 2016 NSF–Sloan Workshop on Practical Privacy Vilhuber, Lars; Schmutte, Ian; Abowd, John M. On October 14, 2016, we hosted a workshop that brought together economists, survey statisticians, and computer scientists with expertise in the field of privacy preserving methods: Census Bureau staff working on implementing cutting-edge methods in the Bureau’s flagship public-use products mingled with academic researchers from a variety of universities. The four products discussed as part of the workshop were 1. the American Community Survey (ACS); 2. Longitudinal Employer-Household Data (LEHD), in particular the LEHD Origin-Destination Employment Statistics (LODES); the 3. 2020 Decennial Census; and the 4. 2017 Economic Census. The goal of the workshop was to 1. Discuss the specific challenges that have arisen in ongoing efforts to apply formal privacy models to Census data products by drawing together expertise of academic and governmental researchers 2. Produce short written memos that summarize concrete suggestions for practical applications to specific Census Bureau priority areas.

**Vilhuber et al.: Proceedings from the 2017 Cornell-Census-NSF-Sloan Workshop on Practical Privacy   ProceedingsNSFSloan2017**

Lars Vilhuber and Ian Schmutte. *Proceedings from the 2017 Cornell-Census-NSF-Sloan Workshop on Practical Privacy*. Document 43. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/43/>.

Abstract: These proceedings report on a workshop hosted at the U.S. Census Bureau on May 8, 2017. Our purpose was to gather experts from various backgrounds together to continue discussing the development of formal privacy systems for Census Bureau data products. This workshop was a successor to a previous workshop held in October 2016 (Vilhuber and Schmutte 2017). At our prior workshop, we hosted computer scientists, survey statisticians, and economists, all of whom were experts in data privacy. At that time we discussed the practical implementation of cutting-edge methods for publishing data with formal, provable privacy guarantees, with a focus on applications to Census Bureau data products. The teams developing those applications were just starting out when our first workshop took place, and we

spent our time brainstorming solutions to the various problems researchers were encountering, or anticipated encountering. For these cutting-edge formal privacy models, there had been very little effort in the academic literature to apply those methods in real-world settings with large, messy data. We therefore brought together an expanded group of specialists from academia and government who could shed light on technical challenges, subject matter challenges and address how data users might react to changes in data availability and publishing standards. In May 2017, we organized a follow-up workshop, which these proceedings report on. We reviewed progress made in four different areas. The four topics discussed as part of the workshop were 1. the 2020 Decennial Census; 2. the American Community Survey (ACS); 3. the 2017 Economic Census; 4. measuring the demand for privacy and for data quality. As in our earlier workshop, our goals were to 1. Discuss the specific challenges that have arisen in ongoing efforts to apply formal privacy models to Census data products by drawing together expertise of academic and governmental researchers; 2. Produce short written memos that summarize concrete suggestions for practical applications to specific Census Bureau priority areas.

**Vilhuber et al.: Proceedings from the 2016 NSF-Sloan Workshop  
on Practical Privacy** **Vilhuber:LDI:2017:33**

---

Lars Vilhuber and Ian M. Schmutte. *Proceedings from the 2016 NSF-Sloan Workshop on Practical Privacy*. Document 33. Labor Dynamics Institute, Cornell University, 2017. URL: <http://digitalcommons.ilr.cornell.edu/ldi/33/>.

Abstract: On October 14, 2016, we hosted a workshop that brought together economists, survey statisticians, and computer scientists with expertise in the field of privacy preserving methods: Census Bureau staff working on implementing cutting-edge methods in the Bureau’s flagship public-use products mingled with academic researchers from a variety of universities. The four products discussed as part of the workshop were 1. the American Community Survey (ACS); 2. Longitudinal Employer-Household Data (LEHD), in particular the LEHD Origin-Destination Employment Statistics (LODES); the 3. 2020 Decennial Census; and the 4. 2017 Economic Census. The goal of the workshop was to 1. Discuss the specific challenges that have arisen in ongoing efforts to apply formal privacy models to Census data products

by drawing together expertise of academic and governmental researchers 2. Produce short written memos that summarize concrete suggestions for practical applications to specific Census Bureau priority areas. Funding for the workshop was provided by the National Science Foundation (CNS-1012593) and the Alfred P. Sloan Foundation. Organizational support was provided by the Research and Methodology Directorate at the U.S. Census Bureau and the Labor Dynamics Institute at Cornell University.

**Weinberg et al.: Effects of a Government-Academic Partnership:  
Has the NSF-Census Bureau Research Network Helped Improve  
the U.S. Statistical System? RePEc:cen:wpaper:17-59r**

---

Daniel H. Weinberg et al. *Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Improve the U.S. Statistical System?* Working Papers 17-59r. Center for Economic Studies, U.S. Census Bureau, 2017. URL: <https://ideas.repec.org/p/cen/wpaper/17-59r.html>.

Abstract: The National Science Foundation-Census Bureau Research Network (NCRN) was established in 2011 to create interdisciplinary research nodes on methodological questions of interest and significance to the broader research community and to the Federal Statistical System (FSS), particularly the Census Bureau. The activities to date have covered both fundamental and applied statistical research and have focused at least in part on the training of current and future generations of researchers in skills of relevance to surveys and alternative measurement of economic units, households, and persons. This paper discusses some of the key research findings of the eight nodes, organized into six topics: (1) Improving census and survey data collection methods; (2) Using alternative sources of data; (3) Protecting privacy and confidentiality by improving disclosure avoidance; (4) Using spatial and spatio-temporal statistical modeling to improve estimates; (5) Assessing data cost and quality tradeoffs; and (6) Combining information from multiple sources. It also reports on collaborations across nodes and with federal agencies, new software developed, and educational activities and outcomes. The paper concludes with an evaluation of the ability of the FSS to apply the NCRN's research outcomes and suggests some next steps, as well as the implications of this research-network model for future federal government renewal initiatives.

**Miranda et al.: Using Partially Synthetic Microdata to Protect Sensitive Cells in Business Statistics**

**miranda-vilhuber-2016-ecommons**

---

Javier Miranda and Lars Vilhuber. *Using Partially Synthetic Microdata to Protect Sensitive Cells in Business Statistics*. Tech. rep. 1813:42339. NSF Census Research Network - NCRN-Cornell, 2016. URL: <http://hdl.handle.net/1813/42339>.

Abstract: We describe and analyze a method that blends records from both observed and synthetic microdata into public-use tabulations on establishment statistics. The resulting tables use synthetic data only in potentially sensitive cells. We describe different algorithms, and present preliminary results when applied to the Census Bureau’s Business Dynamics Statistics and Synthetic Longitudinal Business Database, highlighting accuracy and protection afforded by the method when compared to existing public-use tabulations (with suppressions).

**Miranda et al.: Using Partially Synthetic Microdata to Protect Sensitive Cells in Business Statistics**      **RePEc:cen:wpaper:16-10**

---

Javier Miranda and Lars Vilhuber. *Using Partially Synthetic Microdata to Protect Sensitive Cells in Business Statistics*. Working Papers 16-10. Center for Economic Studies, U.S. Census Bureau, 2016. URL: <https://ideas.repec.org/p/cen/wpaper/16-10.html>.

Abstract: We describe and analyze a method that blends records from both observed and synthetic microdata into public-use tabulations on establishment statistics. The resulting tables use synthetic data only in potentially sensitive cells. We describe different algorithms, and present preliminary results when applied to the Census Bureau’s Business Dynamics Statistics and Synthetic Longitudinal Business Database, highlighting accuracy and protection afforded by the method when compared to existing public-use tabulations (with suppressions).

**Vilhuber et al.: Synthetic Establishment Microdata Around the World**      **vilhuber-abowd-reiter-2016-ecommons**

---

Lars Vilhuber, John A. Abowd, and Jerome P. Reiter. *Synthetic Establishment Microdata Around the World*. Tech. rep. 1813:42340. NSF Census Re-

search Network - NCRN-Cornell, 2016. URL: <http://hdl.handle.net/1813/42340>.

Abstract: In contrast to the many public-use microdata samples available for individual and household data from many statistical agencies around the world, there are virtually no establishment or firm microdata available. In large part, this difficulty in providing access to business micro data is due to the skewed and sparse distributions that characterize business data. Synthetic data are simulated data generated from statistical models. We organized sessions at the 2015 World Statistical Congress and the 2015 Joint Statistical Meetings, highlighting work on synthetic establishment microdata. This overview situates those papers, published in this issue, within the broader literature.

**Jörg Drechsler et al.: A First Step Towards A German SynLBD:  
Constructing A German Longitudinal Business Database**  
**RePEc:cen:wpaper:14-13**

---

Jörg Drechsler and Lars Vilhuber. *A First Step Towards A German SynLBD: Constructing A German Longitudinal Business Database*. Working Papers 14-13. Center for Economic Studies, U.S. Census Bureau, 2014. URL: <http://ideas.repec.org/p/cen/wpaper/14-13.html>.

Abstract: One major criticism against the use of synthetic data has been that the efforts necessary to generate useful synthetic data are so intense that many statistical agencies cannot afford them. We argue many lessons in this evolving field have been learned in the early years of synthetic data generation, and can be used in the development of new synthetic data products, considerably reducing the required investments. The final goal of the project described in this paper will be to evaluate whether synthetic data algorithms developed in the U.S. to generate a synthetic version of the Longitudinal Business Database (LBD) can easily be transferred to generate a similar data product for other countries. We construct a German data product with information comparable to the LBD - the German Longitudinal Business Database (GLBD) - that is generated from different administrative sources at the Institute for Employment Research, Germany. In a future step, the algorithms developed for the synthesis of the LBD will be applied to the GLBD. Extensive evaluations will illustrate whether the algorithms provide useful synthetic data without further adjustment. The ultimate goal



of the project is to provide access to multiple synthetic datasets similar to the SynLBD at Cornell to enable comparative studies between countries. The Synthetic GLBD is a first step towards that goal.

**Miranda et al.: Looking Back On Three Years Of Using The Synthetic LBD Beta**  
**RePEc:cen:wpaper:14-11**

---

Javier Miranda and Lars Vilhuber. *Looking Back On Three Years Of Using The Synthetic LBD Beta*. Working Papers 14-11. Center for Economic Studies, U.S. Census Bureau, 2014. URL: <http://ideas.repec.org/p/cen/wpaper/14-11.html>.

Abstract: Distributions of business data are typically much more skewed than those for household or individual data and public knowledge of the underlying units is greater. As a results, national statistical offices (NSOs) rarely release establishment or firm-level business microdata due to the risk to respondent confidentiality. One potential approach for overcoming these risks is to release synthetic data where the establishment data are simulated from statistical models designed to mimic the distributions of the real underlying microdata. The US Census Bureau's Center for Economic Studies in collaboration with Duke University, the National Institute of Statistical Sciences, and Cornell University made available a synthetic public use file for the Longitudinal Business Database (LBD) comprising more than 20 million records for all business establishment with paid employees dating back to 1976. The resulting product, dubbed the SynLBD, was released in 2010 and is the first-ever comprehensive business microdata set publicly released in the United States including data on establishments employment and payroll, birth and death years, and industrial classification. This paper documents the scope of projects that have requested and used the SynLBD.

**Vilhuber et al.: LEHD Infrastructure files in the Census RDC - Overview**  
**RePEc:cen:wpaper:14-26**

---

Lars Vilhuber and Kevin McKinney. *LEHD Infrastructure files in the Census RDC - Overview*. Working Papers 14-26. Center for Economic Studies, U.S. Census Bureau, 2014. URL: <http://ideas.repec.org/p/cen/wpaper/14-26.html>.



Abstract: The Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau, with the support of several national research agencies, maintains a set of infrastructure files using administrative data provided by state agencies, enhanced with information from other administrative data sources, demographic and economic (business) surveys and censuses. The LEHD Infrastructure Files provide a detailed and comprehensive picture of workers, employers, and their interaction in the U.S. economy. This document describes the structure and content of the 2011 Snapshot of the LEHD Infrastructure files as they are made available in the Census Bureaus secure and restricted-access Research Data Center network. The document attempts to provide a comprehensive description of all researcher-accessible files, of their creation, and of any modifications made to the files to facilitate researcher access.

**Bérubé et al.: Estimation de la contribution de la réallocation de la main-d'œuvre à la croissance de la productivité au Canada**  
**BerubeDostieVilhuber2013**

---

Charles Bérubé, Benoit Dostie, and Lars Vilhuber. *Estimation de la contribution de la réallocation de la main-d'œuvre à la croissance de la productivité au Canada*. Tech. rep. Centre sur la productivité et la prospérité, HEC Montréal, 2013. URL: [http://cpp.hec.ca/cms/assets/documents/recherches\\_publiees/CH\\_2012\\_01.pdf](http://cpp.hec.ca/cms/assets/documents/recherches_publiees/CH_2012_01.pdf).

Abstract: In this report, we estimate the contribution of labour reallocation to productivity growth in the Canadian manufacturing sector. We find that most of productivity growth comes from within firm improvements, leaving a limited role for labour reallocation. Still, we also find that the importance of labour reallocation increase over time. This is both due to increasing net-entry and inter-firm effects. These effects are much more important post 2000 than in the 1990s. We also find that lost production from exiting firms is now most likely replaced by production from existing firms, while previously, it was more likely to be replaced by production from new firms. (French only).

**Vilhuber: Methods for Protecting the Confidentiality of Firm-Level Data: Issues and Solutions**  
**Vilhuber2013**

---

Lars Vilhuber. *Methods for Protecting the Confidentiality of Firm-Level Data:*

*Issues and Solutions*. Document 19. Labor Dynamics Institute, 2013. URL: <http://digitalcommons.ilr.cornell.edu/ldi/19/>.

Abstract: This report will provide an overview of methods used by statistical agencies to encourage, support, and enhance research access to data for the purpose of generating new knowledge. Quite a few reports and scientific articles have addressed the issue before, and we will be highly indebted to that literature. To a summary of that literature, we hope to provide some recent developments and experiences derived from a decade of working with systems that increase access as both researchers as well as data providers. The report will focus on the data provided by statistical agencies, but it should be understood that government agencies other than a National Statistical Office (NSO) may acquire that function. While excluding the legal background limiting or permitting such data collection and provision, we will highlight some alternate sources and methods, prior to concluding.

**John M. Abowd et al.: Dynamically Consistent Noise Infusion and Partially Synthetic Data as Confidentiality Protection Measures for Related Time Series**  
**AbowdEtAl2012**

---

John M. Abowd, Kaj Gittings, Kevin L. McKinney, Bryce E. Stephens, Lars Vilhuber, and Simon Woodcock. *Dynamically Consistent Noise Infusion and Partially Synthetic Data as Confidentiality Protection Measures for Related Time Series*. Research Conference Papers. Federal Committee on Statistical Methodology, 2012. URL: <http://ideas.repec.org/p/cen/wpaper/12-13.html>.

**John M. Abowd et al.: National Quarterly Workforce Indicators, r2254**  
**NQWI**

---

John M. Abowd and Lars Vilhuber. *National Quarterly Workforce Indicators, r2254*. [Computer file]. Ithaca, NY, USA: Cornell University, Labor Dynamics Institute [distributor], 2012. URL: <http://www2.vrdc.cornell.edu/news/data/qwi-national-data/>.

**K. McKinney et al.: LEHD Data Documentation LEHD-OVERVIEW-S2008-rev1**  
**RePEc:cen:wpaper:11-43**

---

Kevin McKinney and Lars Vilhuber. *LEHD Data Documentation LEHD-*

*OVERVIEW-S2008-rev1*. Working Papers 11-43. Center for Economic Studies, U.S. Census Bureau, 2011. URL: <http://ideas.repec.org/p/cen/wpaper/11-43.html>.

Abstract: No abstract is available for this item.

---

**K. McKinney et al.: LEHD Infrastructure Files in the Census RDC: Overview of S2004 Snapshot** **RePEc:cen:wpaper:11-13**

---

Kevin McKinney and Lars Vilhuber. *LEHD Infrastructure Files in the Census RDC: Overview of S2004 Snapshot*. Working Papers 11-13. Center for Economic Studies, U.S. Census Bureau, 2011. URL: <http://ideas.repec.org/p/cen/wpaper/11-13.html>.

Abstract: The Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau, with the support of several national research agencies, has built a set of infrastructure files using administrative data provided by state agencies, enhanced with information from other administrative data sources, demographic and economic (business) surveys and censuses. The LEHD Infrastructure Files provide a detailed and comprehensive picture of workers, employers, and their interaction in the U.S. economy. This document describes the structure and content of the 2004 Snapshot of the LEHD Infrastructure files as they are made available in the Census Bureau's Research Data Center network.

---

**Von Schrader et al.: New York State Disability and Employment Status Report, 2011** **Employment2011**

---

Sarah Von Schrader, William Erickson, Thomas Golden, and Lars Vilhuber. *New York State Disability and Employment Status Report, 2011*. Report on behalf of New York Makes Work Pay Comprehensive Employment System Medicaid Infrastructure Grant. 2011. URL: [http://ilr-edi-r1.ilr.cornell.edu/nymakesworkpay/docs/Report\\_Card\\_2011/NYS%20Report%20Card%202011.pdf](http://ilr-edi-r1.ilr.cornell.edu/nymakesworkpay/docs/Report_Card_2011/NYS%20Report%20Card%202011.pdf) (visited on 04/10/2014).

---

**John M. Abowd et al.: National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail (with color graphs)** **ces-wp-10-11**

---

John M. Abowd and Lars Vilhuber. *National Estimates of Gross Employment*

*and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail (with color graphs)*. Working Papers 10-11. Center for Economic Studies, U.S. Census Bureau, 2010. URL: <http://ideas.repec.org/p/cen/wpaper/10-11.html>.

Abstract: The Quarterly Workforce Indicators (QWI) are local labor market data produced and released every quarter by the United States Census Bureau. Unlike any other local labor market series produced in the U.S. or the rest of the world, the QWI measure employment flows for workers (accession and separations), jobs (creations and destructions) and earnings for demographic subgroups (age and gender), economic industry (NAICS industry groups), detailed geography (block (experimental), county, Core- Based Statistical Area, and Workforce Investment Area), and ownership (private, all) with fully interacted publication tables. The current QWI data cover 47 states, about 98% of the private workforce in those states, and about 92% of all private employment in the entire economy. State participation is sufficiently extensive to permit us to present the first national estimates constructed from these data. We focus on worker, job, and excess (churning) reallocation rates, rather than on levels of the basic variables. This permits comparison to existing series from the Job Openings and Labor Turnover Survey and the Business Employment Dynamics Series from the Bureau of Labor Statistics. The national estimates from the QWI are an important enhancement to existing series because they include demographic and industry detail for both worker and job flow data compiled from underlying micro-data that have been integrated at the job and establishment levels by the Longitudinal Employer-Household Dynamics Program at the Census Bureau. The estimates presented herein were compiled exclusively from public-use data series and are available for download.

---

**Vilhuber: Measuring firm-level displacement events with administrative data** **Vilhuber2010**

---

Lars Vilhuber. *Measuring firm-level displacement events with administrative data*. Tech. rep. Mannheim, Germany: Workshop on Measurement Error in Administrative Data, 2010.

**Von Schrader et al.: New York State Disability and Employment Status Report, 2009**  
**Employment2009**

---

Sarah Von Schrader, William Erickson, Lars Vilhuber, and Thomas. Golden. *New York State Disability and Employment Status Report, 2009*. Report on behalf of New York Makes Work Pay Comprehensive Employment System Medicaid Infrastructure Grant. Cornell University, Employment and Disability Institute, 2010. URL: <http://digitalcommons.ilr.cornell.edu/edicollect/1282/> (visited on 04/10/2014).

**Vilhuber: Adjusting imperfect data: Overview and case studies**  
**Vilhuber2007**

---

Lars Vilhuber. *Adjusting imperfect data: Overview and case studies*. Working paper 12977. NBER, 2007. DOI: [10.3386/w12977](https://doi.org/10.3386/w12977).

File: [tp-2004-05.pdf:L/LEHD/tp-2004-05.pdf:PDF](#).

**John M. Abowd et al.: Confidentiality Protection in the Census Bureau's Quarterly Workforce Indicators**  
**AbowdEtAl2005b**

---

John M. Abowd, Bryce E. Stephens, and Lars Vilhuber. *Confidentiality Protection in the Census Bureau's Quarterly Workforce Indicators*. presented at the Joint Statistical Meetings 2005, Minneapolis, MN. U.S. Census Bureau, LEHD and Cornell University, 2005.

File: [Abowd-Stephens-Vilhuber-2005.pdf:L/LEHD/Abowd-Stephens-Vilhuber-2005.pdf:PDF](#).

**Vilhuber: Adjusting imperfect data: Overview and case studies**  
**tp-2004-05**

---

Lars Vilhuber. *Adjusting imperfect data: Overview and case studies*. Technical paper TP-2004-05. LEHD, 2004.

File: [tp-2004-05.pdf:L/LEHD/tp-2004-05.pdf:PDF](#).

**John M. Abowd et al.: The Creation of the Employment Dynamics Estimates**  
**tp-2002-13**

---

John M. Abowd, Paul A. Lengermann, and Lars Vilhuber. *The Creation of*

*the Employment Dynamics Estimates*. Technical paper TP-2002-13. LEHD, U.S. Census Bureau, 2002. URL: <https://ideas.repec.org/p/cen/tpaper/2002-13.html>.

File: [tp-2002-13.pdf:L/LEHD/tp-2002-13.pdf:PDF](#).

---

**John M. Abowd et al.: The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers** **tp-2002-17**

---

John M. Abowd and Lars Vilhuber. *The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers*. Technical paper TP-2002-17. LEHD, U.S. Census Bureau, 2002.

File: [tp-2002-17.pdf:L/LEHD/tp-2002-17.pdf:PDF](#).

---

**Bowlus et al.: Displaced workers, early leavers, and re-employment wages** **tp-2002-18**

---

Audra Bowlus and Lars Vilhuber. *Displaced workers, early leavers, and re-employment wages*. Technical paper TP-2002-18. LEHD, U.S. Census Bureau, 2002.

File: [tp-2002-18.pdf:L/LEHD/tp-2002-18.pdf:PDF](#).

---

**Lengermann et al.: Abandoning the Sinking Ship: The Composition of Worker Flows Prior to Displacement** **tp-2002-11**

---

Paul A. Lengermann and Lars Vilhuber. *Abandoning the Sinking Ship: The Composition of Worker Flows Prior to Displacement*. Technical paper TP-2002-11. LEHD, U.S. Census Bureau, 2002.

File: [tp-2002-11.pdf:L/LEHD/tp-2002-11.pdf:PDF](#).

---

**H. J. Holzer et al.: Escaping poverty for low-wage workers: The role of employer characteristics and changes** **tp-2001-02**

---

Harry J. Holzer, Julia I. Lane, Lars Vilhuber, Henry Jackson, and George Putnam. *Escaping poverty for low-wage workers: The role of employer characteristics and changes*. Technical paper TP-2001-02. LEHD, U.S. Census Bureau, 2001.

File: [tp-2001-02.pdf:L/LEHD/tp-2001-02.pdf:PDF](#).

**Vilhuber et al.: Longitudinal analysis of SSN response on SIPP  
1990-1993 panels** **tp-2000-01**

---

Lars Vilhuber and Robert Pedace. *Longitudinal analysis of SSN response on SIPP 1990-1993 panels*. Technical paper TP-2000-01. LEHD, U.S. Census Bureau, 2000.

File: [tp-2000-01.pdf:L/LEHD/tp-2000-01.pdf:PDF](#).

**Vilhuber: Continuous Training and sectoral mobility in Germany**  
**Vilhuber99b**

---

Lars Vilhuber. *Continuous Training and sectoral mobility in Germany*. Scientific Series 99s-03. CIRANO, 1999.

Abstract: This article studies mobility patterns of German workers in light of a model of sector-specific human capital. Furthermore, I employ and describe little-used data on continuous on-the-job training occurring after apprenticeships. Results are presented describing the incidence and duration of continuous training. Continuous training is quite common, despite the high incidence of apprenticeships which precedes this part of a worker's career. Most previous studies have only distinguished between firm-specific and general human capital, generally concluding that training was general. Inconsistent with those conclusions, I show that German men are more likely to find a job within the same sector if they have received continuous training in that sector. These results are similar to results obtained for young U.S. workers, and suggest that sector-specific capital is an important feature of very different labor markets. Furthermore, the results suggest that the observed effect of training on mobility is sensitive to the state of the business cycle, indicating a more complex interaction between supply and demand that most theoretical models allow for.

**Vilhuber: Sector-Specific On-The-Job Training: Evidence from U.S.  
Data** **Vilhuber97a**

---

Lars Vilhuber. *Sector-Specific On-The-Job Training: Evidence from U.S. Data*. Scientific Series 97s-42. CIRANO, 1997.

Abstract: Using data from the National Longitudinal Survey of Youth (NLSY), we re-examine the effect of formal on-the-job training on mobility patterns of

young American workers. By employing parametric duration models, we evaluate the economic impact of training on productive time with an employer. Confirming previous studies, we find a positive and statistically significant impact of formal on-the-job training on tenure with the employer providing the training. However, expected duration net of the time spent in the training program is generally not significantly increased. We proceed to document and analyze intra-sectoral and cross-sectoral mobility patterns in order to infer whether training provides firm-specific, industry-specific, or general human capital. The econometric analysis rejects a sequential model of job separation in favor of a competing risks specification. We find significant evidence for the industry-specificity of training. The probability of sectoral mobility upon job separation decreases with training received in the current industry, whether with the last employer or previous employers, and employment attachment increases with on-the-job training. These results are robust to a number of variations on the base model.

### **Vilhuber: Wage Flexibility and Contract Structure in Germany** **Vilhuber96**

---

Lars Vilhuber. *Wage Flexibility and Contract Structure in Germany*. Scientific Series 96s-28. CIRANO, 1996.

Abstract: In this paper, we look at how labor market conditions at different points during the tenure of individuals with firms are correlated with current earnings. Using data from the German Socioeconomic Panel on individuals for the period 1984 to 1994, we find that both the contemporaneous unemployment rate and prior values of the unemployment rate are significantly correlated with current earnings, contrary to results for the American labor market. We interpret this result as evidence that German unions do in fact bargain over both wages and employment, but that the models of individualistic contracts, such as the implicit contract model, may explain some of the observed wage drift and longer-term wage movements reasonably well. Furthermore, we explore the heterogeneity of contracts over a variety of worker and job characteristics. In particular, we find evidence that contracts differ across industries and across firm size. Workers of large firms are remarkably more insulated from the job market than workers for any other type of firm, indicating the importance of internal job markets.



## 5 Online resources

---

**Perry et al.: CED<sup>2</sup>AR: Comprehensive Extensible Data Documentation and Access Repository**

---

**ced2ar**

Benjamin Perry, Jeremy Williams, Lars Vilhuber, and William Block. *CED<sup>2</sup>AR: Comprehensive Extensible Data Documentation and Access Repository*. Cornell University, for NSF Grant SES-1131848. 2013. URL: <http://www2.ncrn.cornell.edu/ced2ar-web/> (visited on 04/10/2014).

---

**Karr et al.: NSF-Census Research Network**

---

**ncrn.info**

Alan Karr, Lars Vilhuber, Jamie Nunnally, and Katherine Kantner. *NSF-Census Research Network*. National Institute for the Statistical Sciences (NISS), Cornell University, and Duke University, for NSF Grant SES-1237602. 2012. URL: <http://www.ncrn.info> (visited on 04/10/2014).

---

**Vilhuber et al.: NSF-Census Research Network - Cornell node website**

---

**ncrn.cornell**

Lars Vilhuber, Benjamin Perry, William Block, and Jeremy Williams. *NSF-Census Research Network - Cornell node website*. Cornell University, for NSF Grant SES-1131848. 2012. URL: <http://www.ncrn.cornell.edu> (visited on 04/10/2014).

---

**John M. Abowd et al.: VirtualRDC - Synthetic Data Server**

---

**AbowdVilhuber**

John M. Abowd and Lars Vilhuber. *VirtualRDC - Synthetic Data Server*. Cornell University, Labor Dynamics Institute. 2010. URL: <http://www.vrdc.cornell.edu/sds/>.

---

**Von Schrader et al.: New York State Disability and Employment Status Report, 2010**

---

**Employment2010**

Sarah Von Schrader, William Erickson, Thomas Golden, and Lars Vilhuber. *New York State Disability and Employment Status Report, 2010*. Cornell University, Employment and Disability Institute on behalf of New York Makes Work Pay Comprehensive Employment System Medicaid Infrastruc-

ture Grant. 2010. URL: <http://www.nymakesworkpay.org/status-reports/index.cfm> (visited on 04/10/2014).

**Von Schrader et al.: County-level Disability and Employment Status Reports, 2007** **Employment2007**

---

Sarah Von Schrader, William Erickson, Lars Vilhuber, and Thomas Golden. *County-level Disability and Employment Status Reports, 2007*. Cornell University, Employment and Disability Institute on behalf of New York Makes Work Pay Comprehensive Employment System Medicaid Infrastructure Grant. 2009. URL: [http://www.ilr.cornell.edu/edi/nymakesworkpay/policy/stats\\_2009.cfm](http://www.ilr.cornell.edu/edi/nymakesworkpay/policy/stats_2009.cfm) (visited on 07/01/2009).

**Von Schrader et al.: County-level Disability and Employment Status Reports, 2009** **Employment2009online**

---

Sarah Von Schrader, William Erickson, Lars Vilhuber, and Thomas. Golden. *County-level Disability and Employment Status Reports, 2009*. Cornell University, Employment and Disability Institute on behalf of New York Makes Work Pay Comprehensive Employment System Medicaid Infrastructure Grant. 2009. URL: <http://www.ilr.cornell.edu/edi/nymakesworkpay/policy/index.cfm> (visited on 01/01/2010).

**John M. Abowd et al.: VirtualRDC** **vrdc**

---

John M. Abowd and Lars Vilhuber. *VirtualRDC*. Cornell University, Labor Dynamics Institute. 2004. URL: <http://www.vrdc.cornell.edu/>.