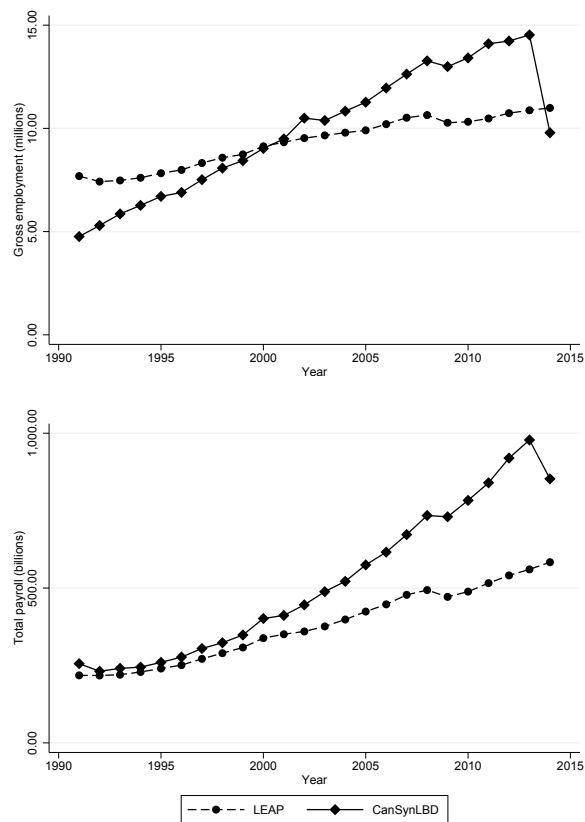


Applying Data Synthesis for Longitudinal Business Data across Three Countries

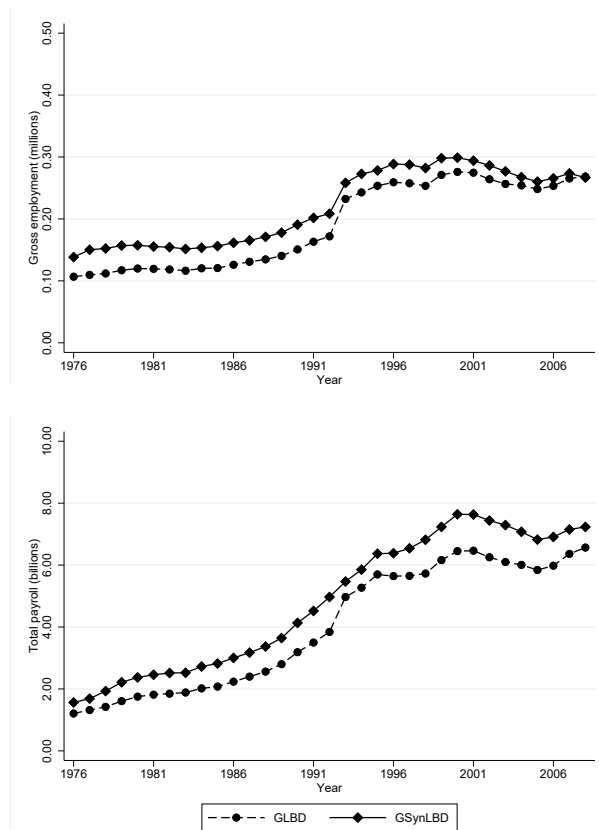
M. Jahangir Alam, Benoit Dostie, Jörg Drechsler, Lars Vilhuber*

February 15, 2020

*Corresponding author: Lars Vilhuber, lars.vilhuber@cornell.edu. Drechsler is an employee of the Institute for Employment Research, IAB. The opinions expressed here are those of the authors, and do not reflect the opinions of any of the statistical agencies involved. All results were reviewed for disclosure risks by their respective custodians, and released to the authors. Alam thanks Claudiu Motoc and Danny Leung for help with the Canadian data. Vilhuber acknowledges funding through NSF Grants SES-1131848 and SES-1042181, and a grant from Alfred P. Sloan Grant (G-2015-13903). Alam and Dostie acknowledge funding through SSHRC Partnership Grant “Productivity, Firms and Incomes”. The creation of the Synthetic LBD was funded by NSF Grant SES-0427889.

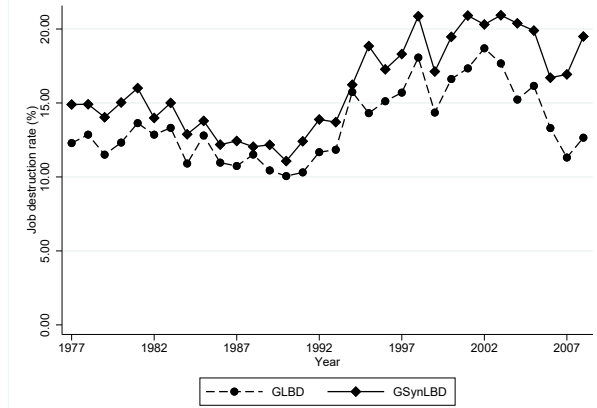
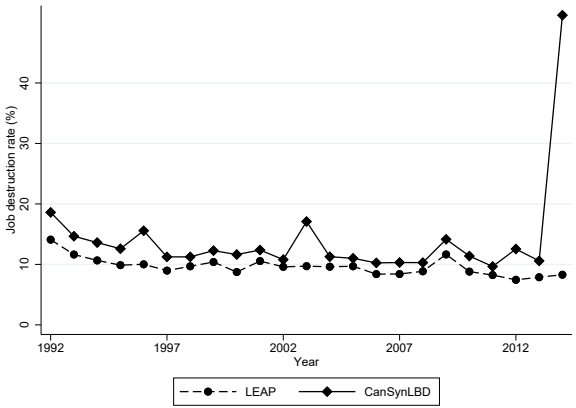
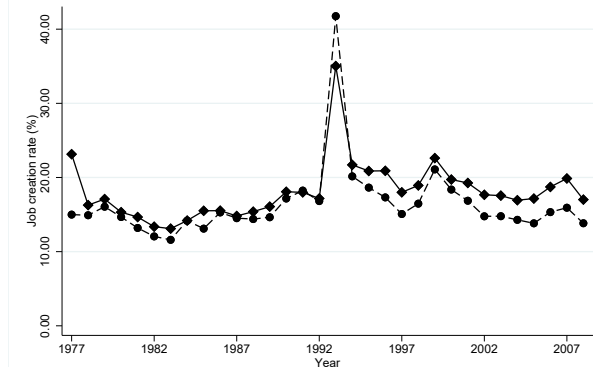
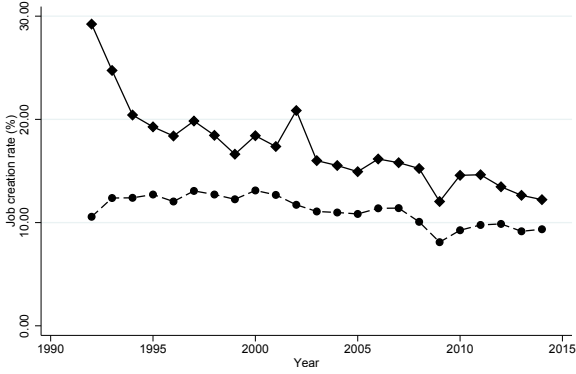


(a) CanSynLBD



(b) GSynLBD

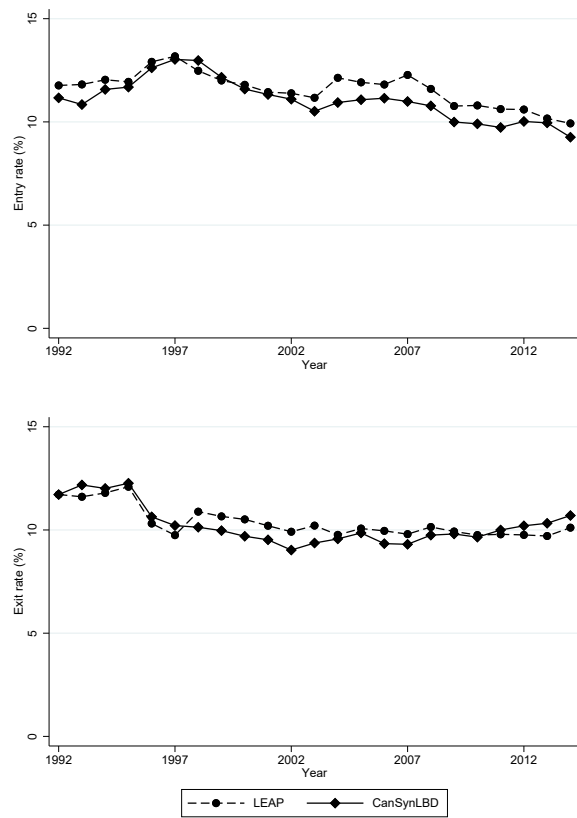
Figure 1: Gross employment level (upper panels) and total payroll (lower panels) by year.



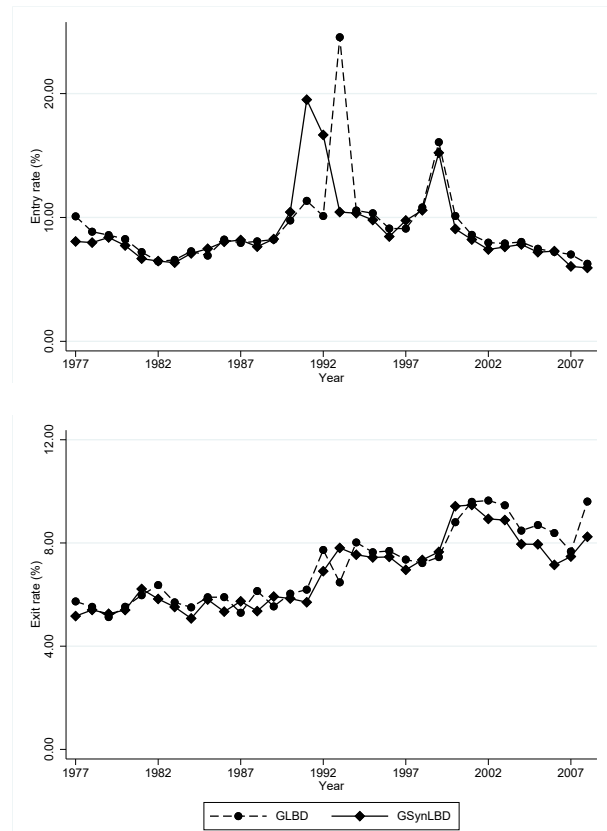
(a) CanSynLBD

(b) GSynLBD

Figure 2: Job creation rates (upper panels) and job destruction rates (lower panels) by year.

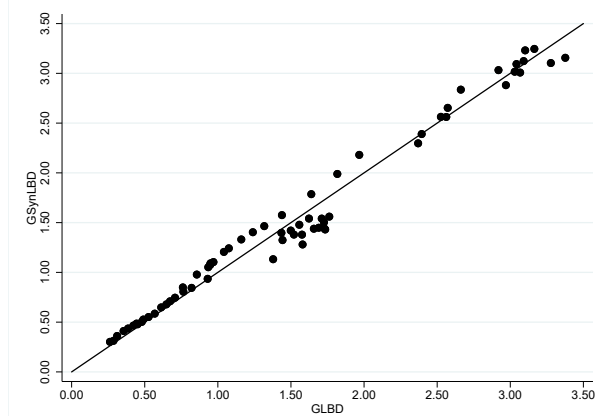
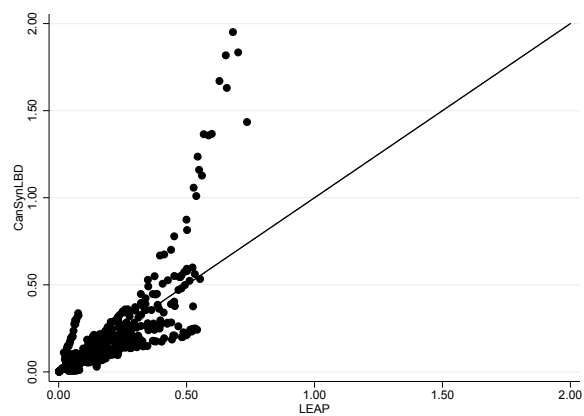
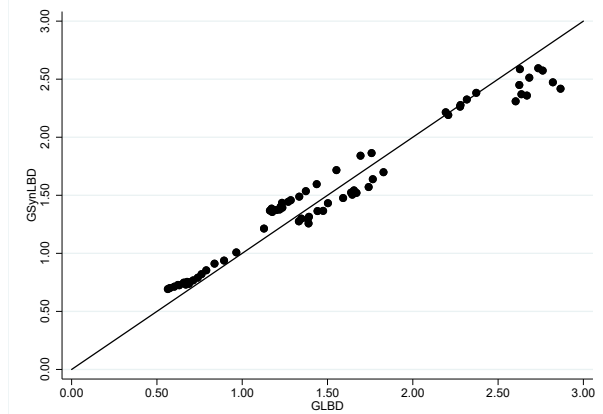
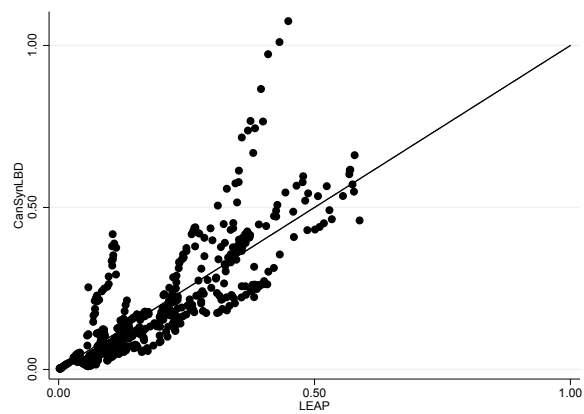
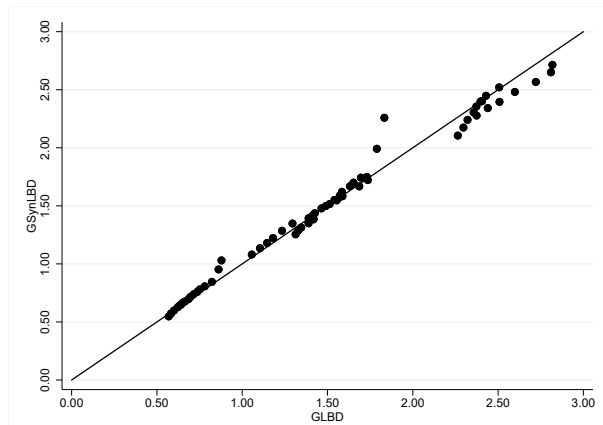
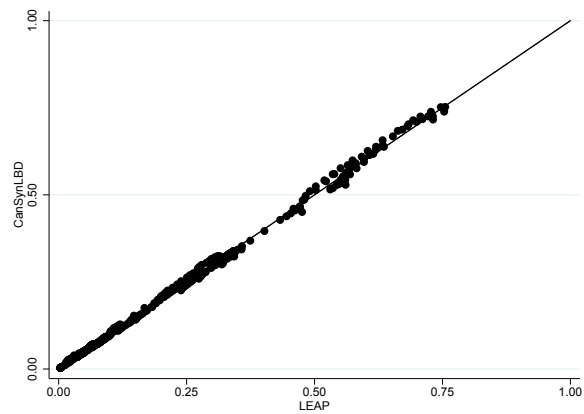


(a) CanSynLBD



(b) GSynLBD

Figure 3: Entry rates (upper panels) and exit rates (lower panels) by year.



(a) CanSynLBD

(b) GSynLBD

Figure 4: Share of entities (upper panels), share of employment (middle panels), and share of payroll (lower panels) by year and industry.

Table 1: m2 and Sargan tests by country

Model	Test	Confidential (CDN)	Synthetic (CDN)	Confidential (GER)	Synthetic (GER)
GMM	m2	-14.5	-27.54	-2.51	-4.13
GMM	Sargan test	69000	15000	3600	2000
System GMM	m2	-11.43	-41.6	19.49	-8.83
System GMM	Sargan test	77000	18000	4500	2800
System GMM MA	m2	8.2	-40.03	19.03	-11.69
System GMM MA	Sargan test	28000	17000	3100	2500

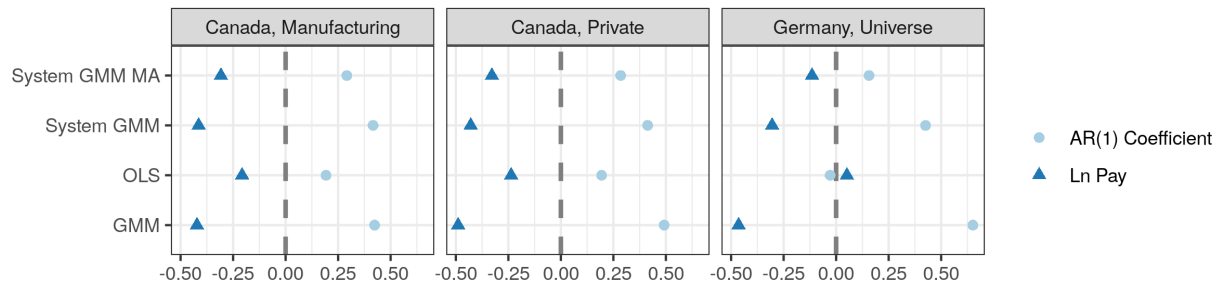


Figure 5: Bias in estimates of coefficients on pay and lagged employment
For details on the estimated coefficients, see the Appendix.