UNIVERSITÉ LAVAL
Faculté des sciences sociales
Département d'économique

Chaire de recherche Industrielle Alliance
sur les enjeux économiques
des changements démographiques

Les données massives et les politiques publiques

Guy Lacroix

Directeur, département d'économique, Université Laval
Titulaire, Chaire Industrielle Alliance sur les enjeux économiques
des changements démographiques

Le mercredi 18 mai 2016

Plan de la présentation

- L'évaluation des politiques en perspective
- Données massives ?
- Tendances lourdes en sciences sociales
- Les enjeux
- La richesse des données massives pour les économistes
- Conclusion

Un peu d'histoire

Jeune étudiant :

- Systèmes centraux
- Rubans magnétiques (enquêtes, 160 mégaoctets)
- Cartes perforées

Un peu d'histoire

Jeune étudiant :

- Systèmes centraux
- Rubans magnétiques (enquêtes, 160 mégaoctets)
- Cartes perforées

Jeune professeur :

- Systèmes centraux
- Rubans magnétiques (enquêtes, 160 mégaoctets)
- Terminaux, systèmes VM/CMS IBM 370
- Décentralisation du calcul
- Inititative de démocratisation des données (IDD)

Un peu d'histoire

Jeune professeur :

- Systèmes centraux
- Rubans magnétiques (enquêtes, 160 mégaoctets)
- Terminaux, systèmes VM/CMS IBM 370
- Décentralisation du calcul
- Inititative de démocratisation des données (IDD)

Vieux (moins jeune) professeur :

- Systèmes centraux (back to the future)
- Terminaux, grappes de calcul (Colosse, MPI, SMP, etc.)
- Rubans magnétiques : 185 téraoctets (1 million les anciens)
- On estime qu'il y a 400 exaoctets d'informations stockés sur les rubans
- Centralisation du calcul
- Centres de données de recherche

Domaine privé

- Walmart, Amazone, Google, eBay, Nielsen (code-barres),*etc.*
- mégaoctet (10^6), gigaoctet (10^9), téraoctet (10^{12}), pétaoctet (10^{15}), exaoctet (10^{18}), zettaoctet (10^{21}), yottaoctet (10^{24})
- Wal-Mart : 1 million de transactions/heure. Stockées dans une base de données de 2,5 petaoctets
- Banque de données Nielsen (depuis 2006) :
  - Prix et quantités hebdomadaires sur 2,6 millions code-barres : nourriture, autres bien d'épicerie, santé et hygiène, et plus de 1,100 produits
  - Marque et format
  - Géographie : données scanner data sur 35 000 épiceries, pharmacies, grands détaillants. Couvre 50% du volume d'achat en épicerie aux USA et Canada et 30% des ventes totales.

Domaine public : Données administratives

- Impôts des particuliers (RQAP)
- Assurance-emploi
- Aide-sociale
- RAMQ, Med-Echo
- Charlemagne (Education, ensemble des dossiers des élèves du Québec, maternelle–secondaire, formation aux adultes et la formation professionnelle)

Qu'entend-on par "données massives"?

Domaine public : Données administratives

- Impôts des particuliers (RQAP)
- Assurance-emploi
- Aide-sociale
- RAMQ, Med-Echo
- Charlemagne (Education, ensemble des dossiers des élèves du Québec, maternelle–secondaire, formation aux adultes et la formation professionnelle)

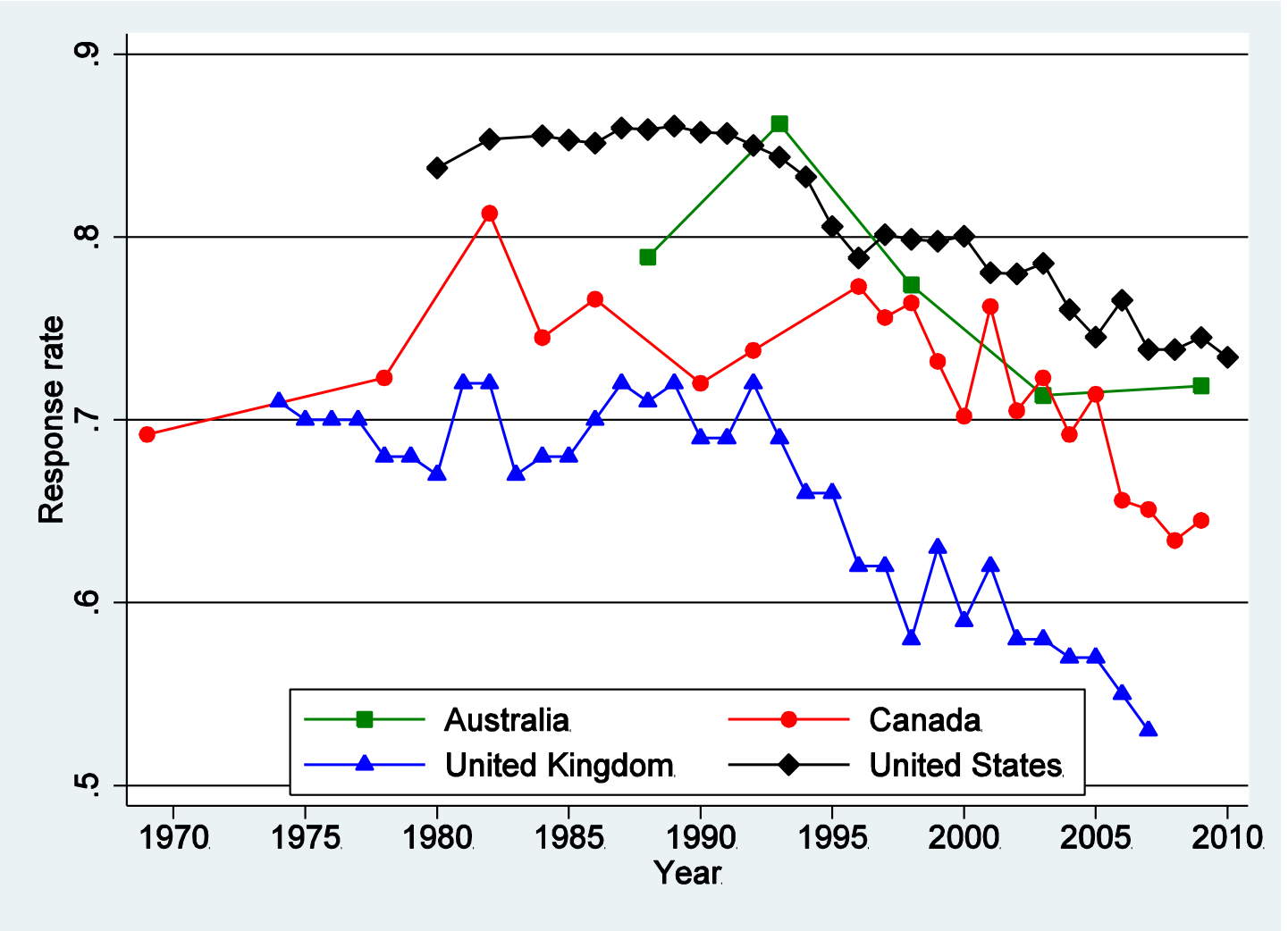
Potentiel de recherche énorme !

- Dynamique de la pauvreté, effets des politiques de réinsertion en emploi, réformes scolaires, réformes en santé, *etc.*

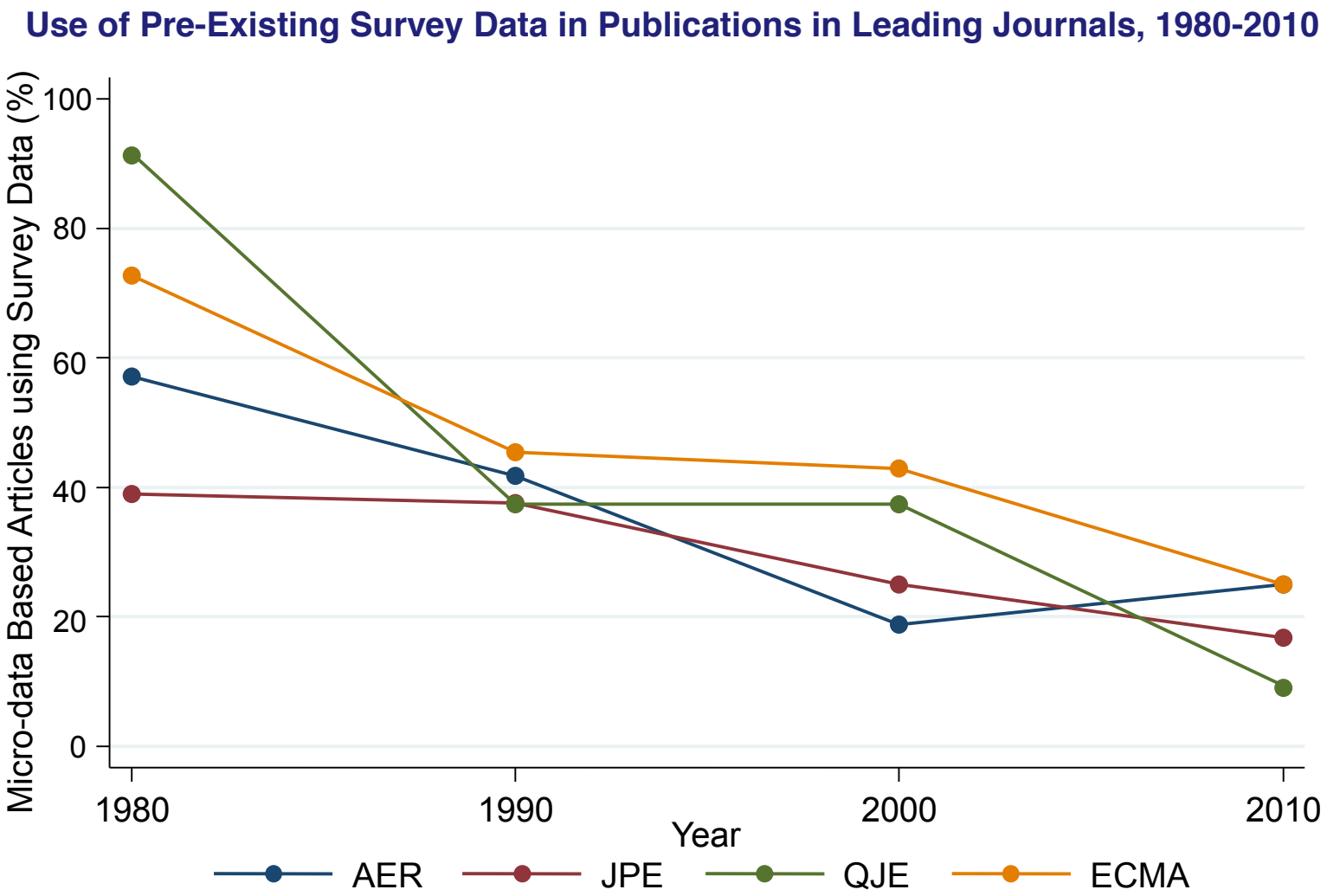
Quatre tendances lourdes en sciences sociales

1. Taux de réponses aux sondages en baisse partout
2. Financement des enquêtes est en baisse
3. Changements technologiques
   - Coûts de stockage des données
   - Coûts de traitement des données
   - Temps de calcul en baisse exponentielle
4. Innovations méthodologiques

Taux de réponses aux sondages en baisse partout

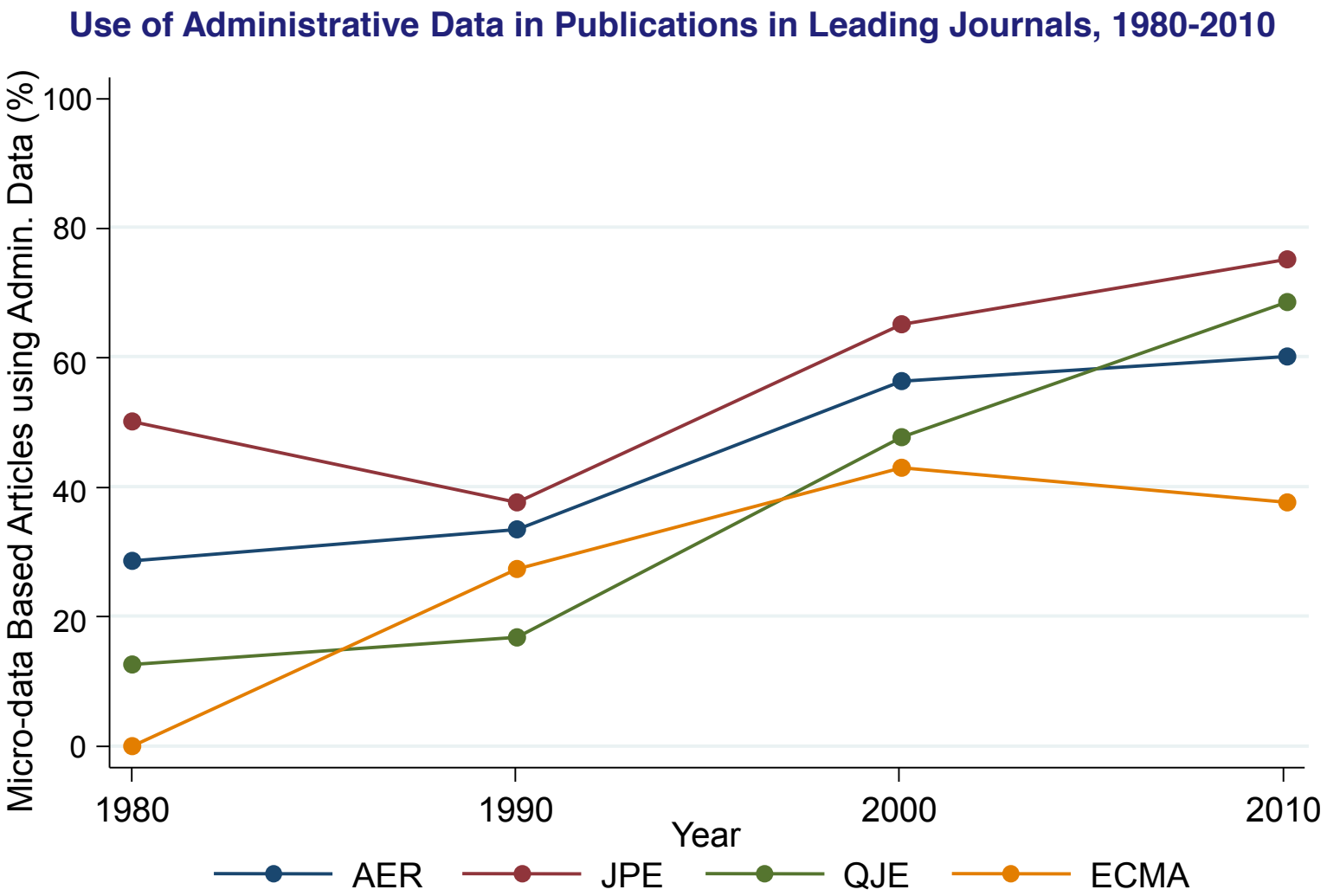


Tendances dans les publications académiques - 1



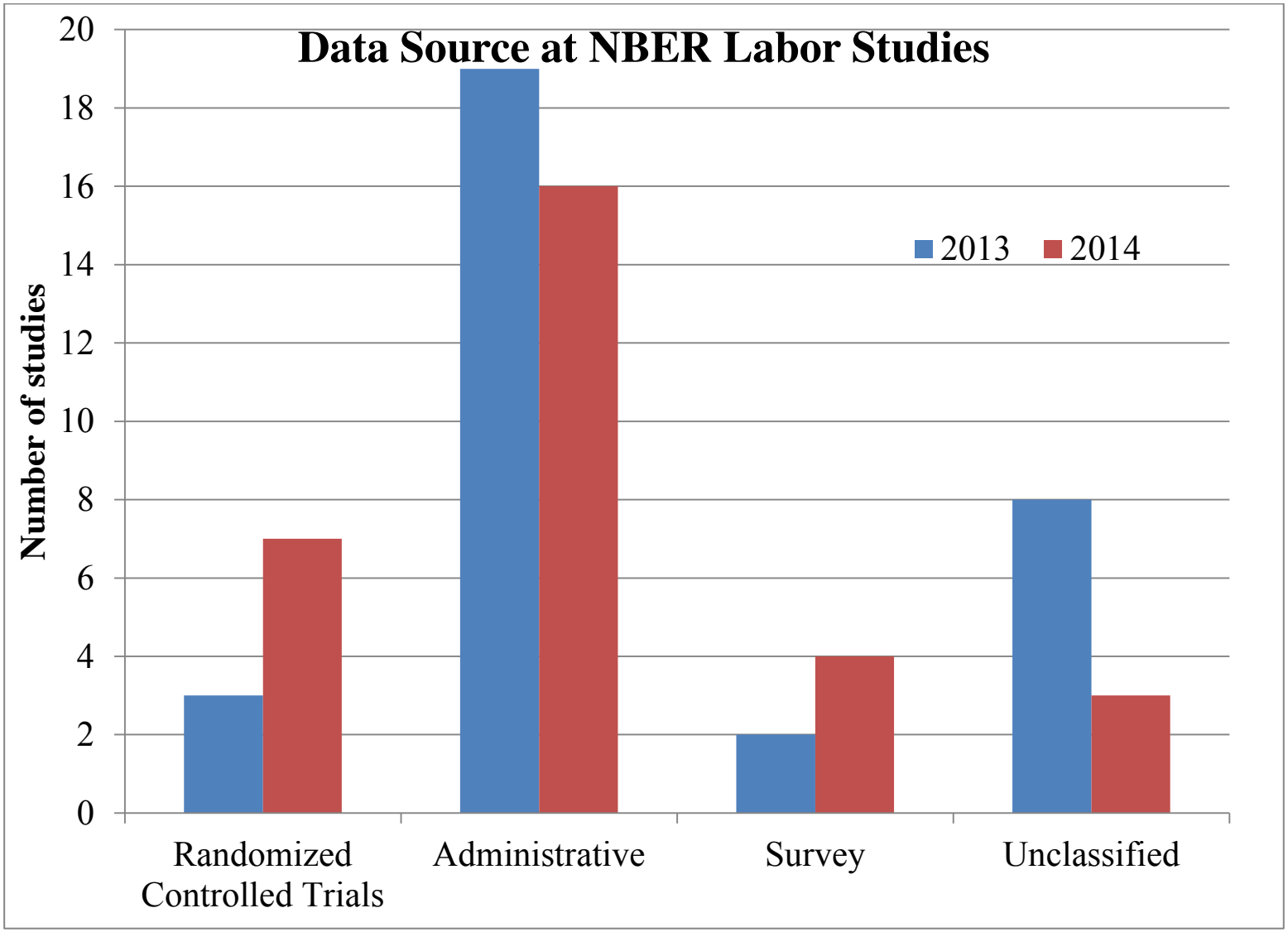
Note: “Pre-existing survey” datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

Tendances dans les publications académiques - 2

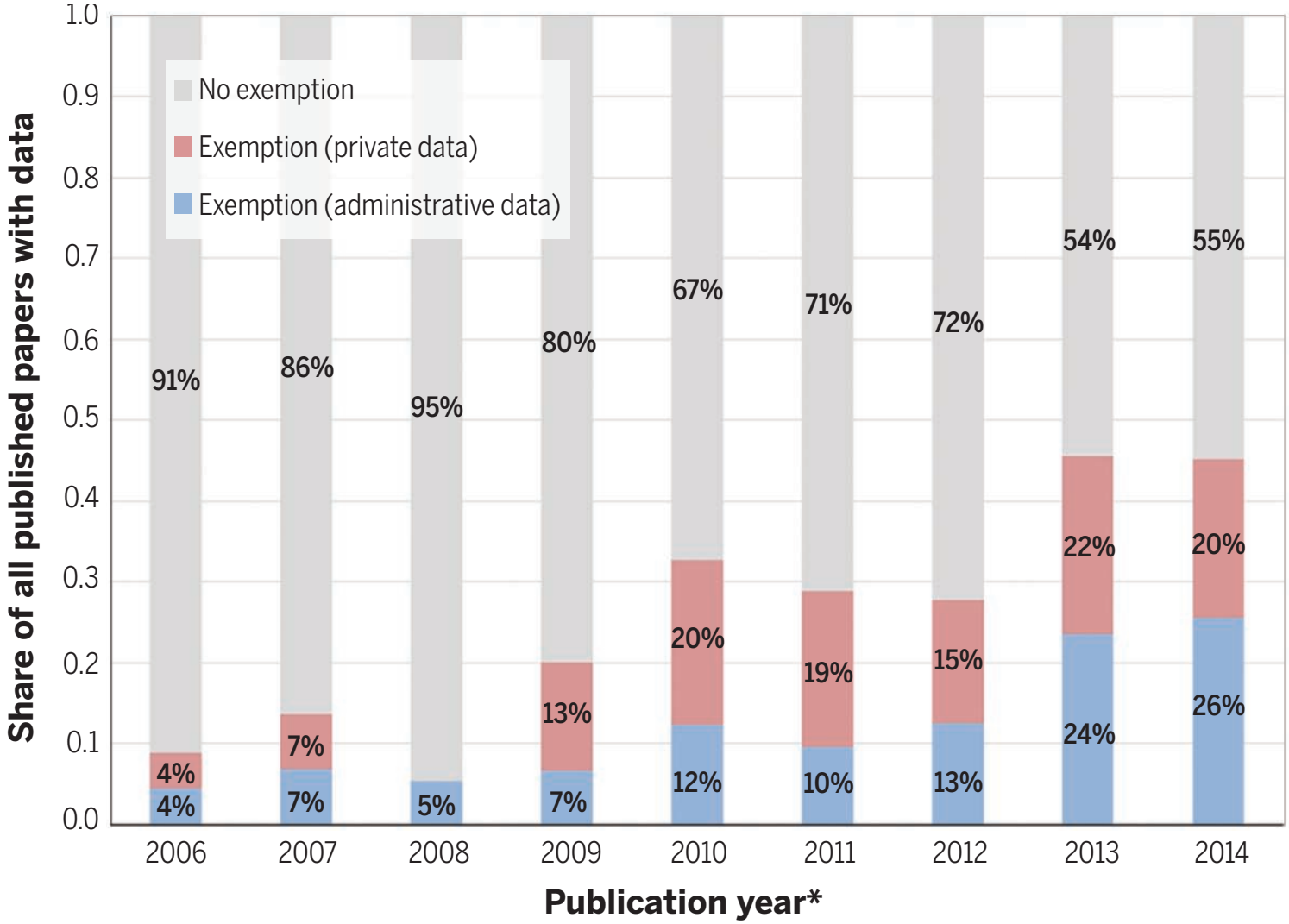


Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

Tendances dans les publications académiques - 3



Tendances dans les publications académiques - 4



Avantages des données administratives

1. Information de meilleure qualité : peu d'attrition ou de valeurs manquantes :
   - CPS (Current Population Survey) : Taux de non-réponse $\approx 31\%$ pour le revenu
   - SIPP (Survey of Income and Program Participation) : Taux d'attrition $\approx 30\%$ sur 3 ans

Avantages des données administratives

1. Information de meilleure qualité : peu d'attrition ou de valeurs manquantes :
   - CPS (Current Population Survey) : Taux de non-réponse $\approx 31\%$ pour le revenu
   - SIPP (Survey of Income and Program Participation) : Taux d'attrition $\approx 30\%$ sur 3 ans
2. Suivi longitudinal sur de longues périodes
   - Taux d'appariement de 95% sur plus de 20 ans dans une étude sur l'impact à long terme de la scolarité dans la petite enfance (Chetty *et al* (2011))

Avantages des données administratives

1. Information de meilleure qualité : peu d'attrition ou de valeurs manquantes :
   - CPS (Current Population Survey) : Taux de non-réponse $\approx 31\%$ pour le revenu
   - SIPP (Survey of Income and Program Participation) : Taux d'attrition $\approx 30\%$ sur 3 ans
2. Suivi longitudinal sur de longues périodes
   - Taux d'appariement de 95% sur plus de 20 ans dans une étude sur l'impact à long terme de la scolarité dans la petite enfance (Chetty *et al* (2011))
3. Très grands échantillons : 2 000 fois la taille du CPS
   - Permet de faire appel à de nouvelles techniques non-paramétriques, de traiter convenablement de l'hétérogénéité des effets des politiques

Innovations méthodologiques

Prédictions hors-échantillon

1. Classifications et "Arbres de régressions" (Regression trees)
2. "Boosting, Bagging, Bootstrap"
3. Random Forest

Sélections de variables

1. LASSO : Least absolute shrinkage and selection operator.
2. Spike and Slab regressions

Économétrie et estimation

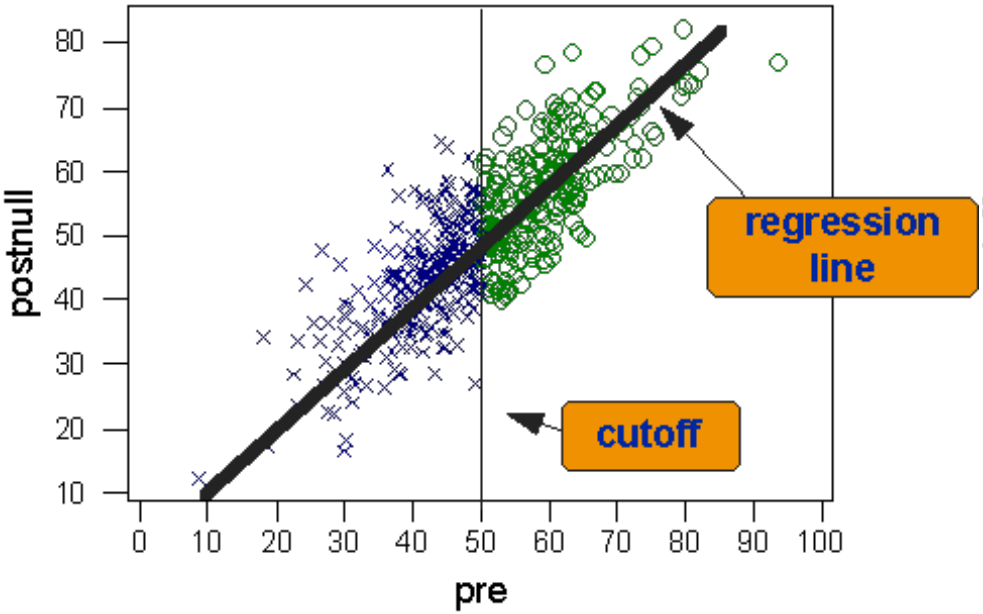
1. Machine Learning
2. Regression Discontinuity

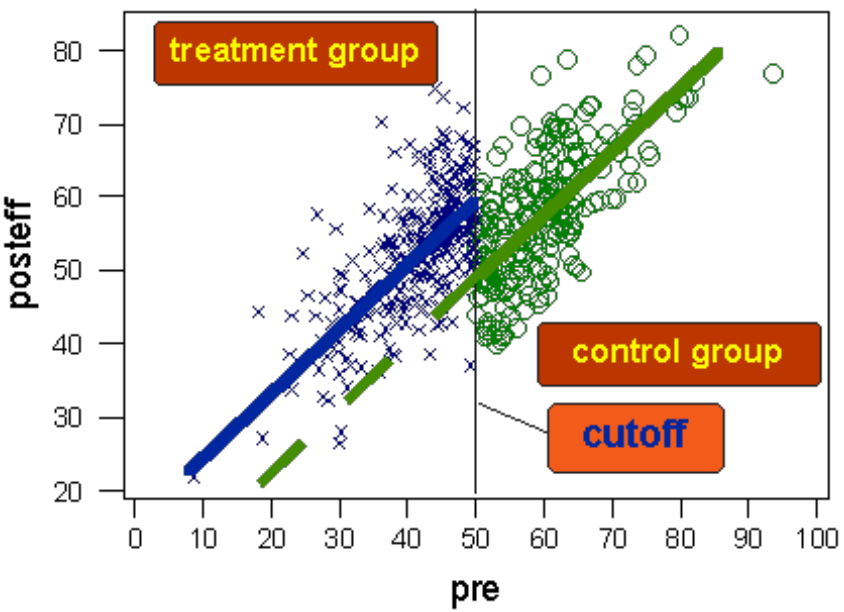
Regression Discontinuity

Méthode statistique qui permet d'évaluer l'effet causal d'une variable (ou d'un traitement) sur une autre en regardant ce qui se passe autour d'un seuil de discontinuité. La méthode s'apparente à ce que les économistes appellent une expérience naturelle.

Regression Discontinuity

Méthode statistique qui permet d'évaluer l'effet causal d'une variable (ou d'un traitement) sur une autre en regardant ce qui se passe autour d'un seuil de discontinuité. La méthode s'apparente à ce que les économistes appellent une expérience naturelle.





Billion Prices Project

Article : Alberto Cavallo and Roberto Rigobon, The Billion Prices Project : Using Online Prices for Measurement and Research, Journal of Economic Perspectives, Volume 30, Number 2, 2016, 151–178

Billion Prices Project

Article : Alberto Cavallo and Roberto Rigobon, The Billion Prices Project : Using Online Prices for Measurement and Research, Journal of Economic Perspectives, Volume 30, Number 2, 2016, 151–178

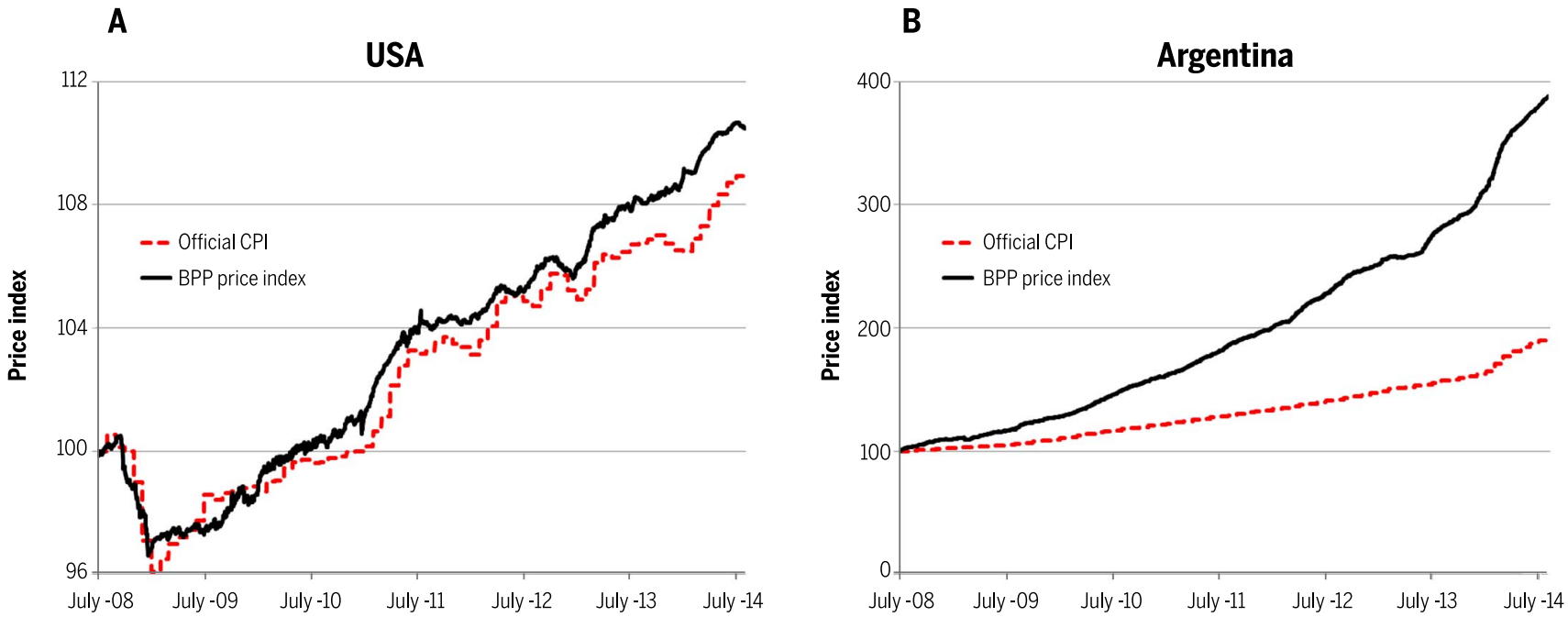
Objectif

- Le gouvernement argentin publiait des taux d'inflation de 8% entre 2007-2011. Les données de prix en ligne indiquaient plutôt 20%, conformément aux estimations des économistes et des gouvernements locaux à partir d'enquêtes
- Développement d'un indice des prix calculé automatiquement, en temps réel à partir des prix en ligne
- Les auteurs ont lancé le BDPP à MIT en 2008.
- 2010 : 5 millions de prix/jour auprès de 300 détaillants dans 50 pays. 500 000 prix aux USA quotidiennement (le Bureau of Labor Statistics collecte environ 80 000 prix par mois ou au 2 mois)

Pourquoi est-ce important pour les politiques publiques ?

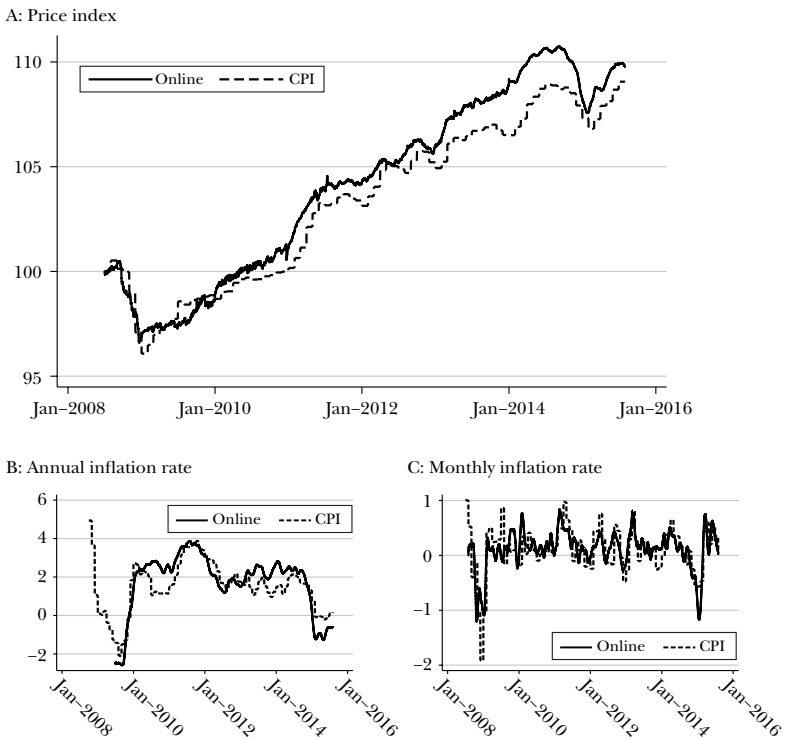
- Rapport Boskin : "Toward A More Accurate Measure Of The Cost Of Living", Final Report to the Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index, December 4, 1996

Billion Prices Project

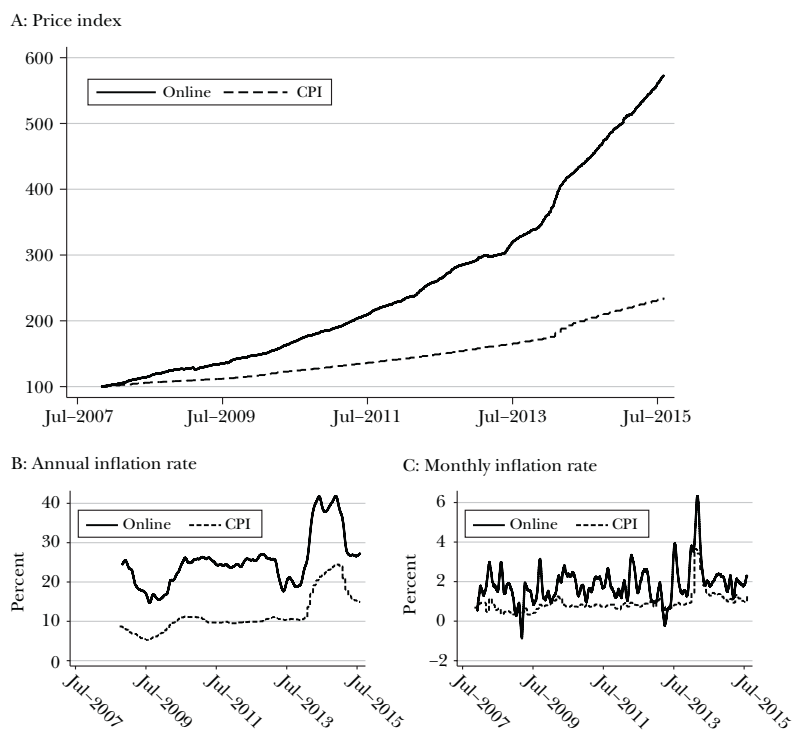


Billion Prices Project

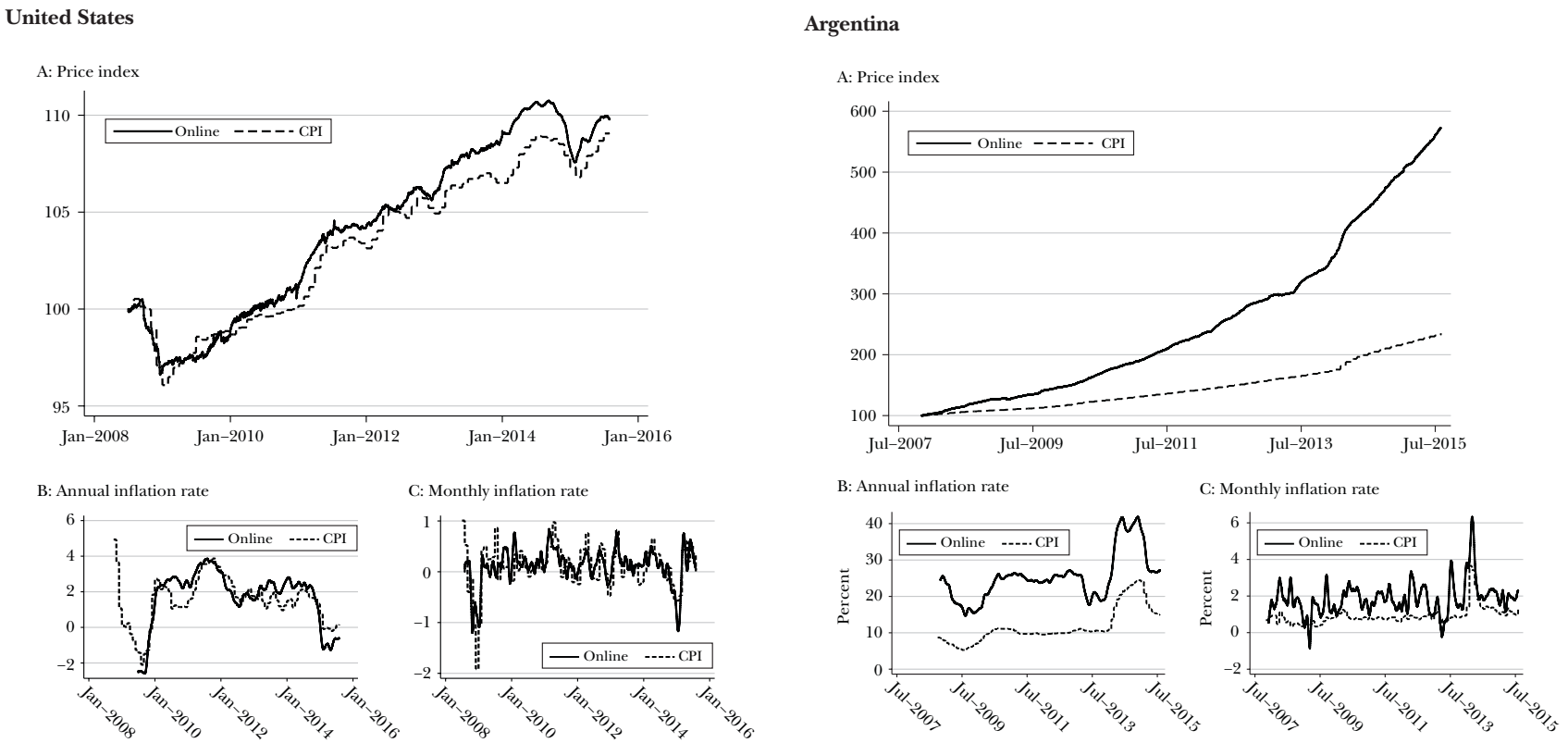
United States



Argentina



Billion Prices Project



Access to the Billion Prices Project Data

As an academic project, we share as much data and results as possible on our webpage (bpp.mit.edu). Most of the micro data and indexes used in our papers are currently available to download on that page, together with detailed scripts that allow others to replicate and extend our results. The micro data are posted with little pretreatment, so other researchers can apply their own methods. We

Mortalité infantile

Article : D. Almond, J. J. Doyle, A. E. Kowalski, H. Williams, Estimating the Marginal Returns to Medical Care : Evidence From At-Risk Newborns, The Quarterly Journal of Economics, May 2010

Mortalité infantile

Article : D. Almond, J. J. Doyle, A. E. Kowalski, H. Williams, Estimating the Marginal Returns to Medical Care : Evidence From At-Risk Newborns, The Quarterly Journal of Economics, May 2010

Objectif :

- "A key policy question is whether the benefits of additional medical expenditures exceed their costs"
- Comparaison de l'effet de traitements différenciés sur la santé des nouveaux-nés dont le poids est de 1 500 grammes (mortalité)

Mortalité infantile

Données :

- "NCHS birth cohort-linked death files", incluent des données sur toutes les naissances aux USA pour les années 1983–1991 et 1995–2002, approximativement 66 million de naissances
- The California research database is the same data set used in Almond and Doyle (2008). These data were collected by the California Office of Statewide Health Planning and Development and include all live births in California from 1991 to 2002 – approximately 6 million births
- Healthcare Cost and Utilization Project (HCUP) : données pour l'Arizona 2001–2006, New Jersey 1995–2006, Maryland 1995–2006, et New York 1995–2000. Approximativement 10,5 million naissances

Données :

- "NCHS birth cohort-linked death files", incluent des données sur toutes les naissances aux USA pour les années 1983–1991 et 1995–2002, approximativement 66 million de naissances
- The California research database is the same data set used in Almond and Doyle (2008). These data were collected by the California Office of Statewide Health Planning and Development and include all live births in California from 1991 to 2002 – approximately 6 million births
- Healthcare Cost and Utilization Project (HCUP) : données pour l'Arizona 2001–2006, New Jersey 1995–2006, Maryland 1995–2006, et New York 1995–2000. Approximativement 10,5 million naissances

Méthode : Regression Discontinuity

Transmission intergénérationnelle de la pauvreté : 1992

Article : D. Zimmerman, "Regression Toward Mediocrity in Economic Stature", The American Economic Review, Vol. 82, No. 3 (Jun., 1992), pp. 409–429

"Few sons, indeed are like their fathers. Generally they are worse ; but just a few are better." [Homer, Odyssey]

"Like father, like son." [Old Proverb]

Transmission intergénérationnelle de la pauvreté : 1992

Article : D. Zimmerman, "Regression Toward Mediocrity in Economic Stature", The American Economic Review, Vol. 82, No. 3 (Jun., 1992), pp. 409–429

"Few sons, indeed are like their fathers. Generally they are worse ; but just a few are better." [Homer, Odyssey]

"Like father, like son." [Old Proverb]

Objectif :

- L'article s'intéresse à la corrélation entre les revenus des pères et de leurs fils sur le cycle de vie
- L'étude trouve que la corrélation intergénérationelle du revenu est d'environ 0,4. Cela suggère que la mobilité intergénérationnelle est beaucoup plus faible que ce que l'on croyait à l'époque

Transmission intergénérationnelle de la pauvreté : 1992

Objectif :

- L'article s'intéresse à la corrélation entre les revenus des pères et de leurs fils sur le cycle de vie
- L'étude trouve que la corrélation intergénérationelle du revenu est d'environ 0,4. Cela suggère que la mobilité intergénérationnelle est beaucoup plus faible que ce que l'on croyait à l'époque

Données :

- National Longitudinal Survey (NLS), "Mature Men" et "Young Men". Enquête débutée en 1966 et comporte quatre cohortes, chacune disposant d'environ 5 000 individus. Seulement 876 dyades pères-fils ont pu être constituées.

Transmission intergénérationnelle de la pauvreté : 2016

Article : R. Chetty, N. Hendren, P. Kline, E. Saez, and N. Turner, "Is the United States Still a Land of Opportunity ? Recent Trends in Intergenerational Mobility", *American Economic Review Papers and Proceedings* 104(5) : 141-147, 2014

Transmission intergénérationnelle de la pauvreté : 2016

Article : R. Chetty, N. Hendren, P. Kline, E. Saez, and N. Turner, "Is the United States Still a Land of Opportunity ? Recent Trends in Intergenerational Mobility", *American Economic Review Papers and Proceedings* 104(5) : 141-147, 2014

Objectif :

- "United States traditionally hailed as 'land of opportunity' "
- Aime à penser que les chances de succès sont indépendantes du revenu des parents
- Controverse à ce sujet en raison des données utilisées (dimension, qualité)

Transmission intergénérationnelle de la pauvreté : 2016

Objectif :

- "United States traditionally hailed as 'land of opportunity' "
- Aime à penser que les chances de succès sont indépendantes du revenu des parents
- Controverse à ce sujet en raison des données utilisées (dimension, qualité)

Données :

- IRS Databank [Chetty, Friedman, Hilger, Saez, Yagan 2011]. Mobilité intergénérationnelle fondée sur des données administratives regroupant plus de 40 millions d'enfants-parents, 1992–2006
- Résultats : Portrait nuancé de la mobilité intergénérationnelle
  - Beaucoup de variation spatiale
  - Régions avec beaucoup de mobilité, d'autres avec pauvreté persistante

Highest Absolute Mobility In The 50 Largest CZs

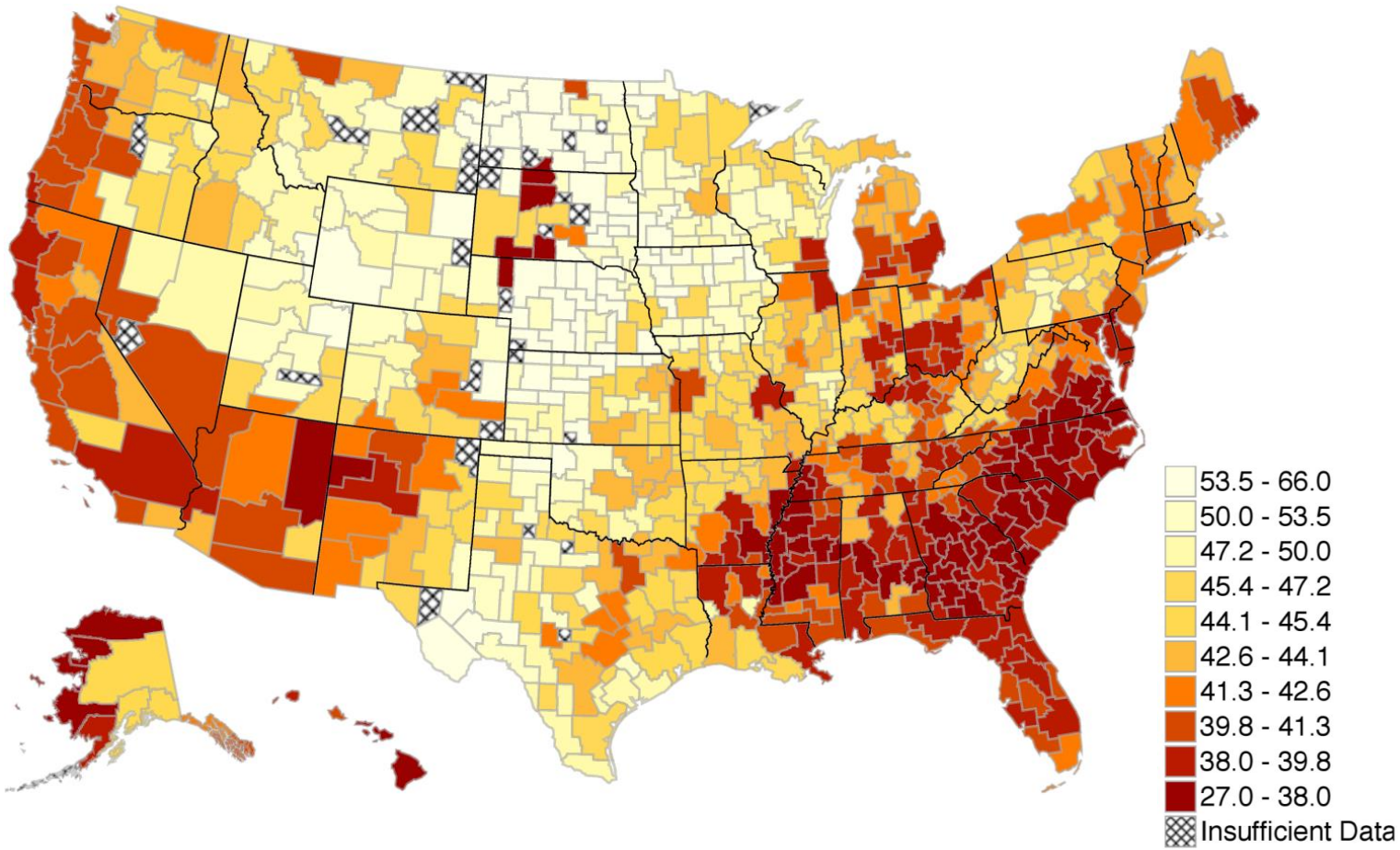
Upward Mobility Rank	CZ Name	Y_{25}	$Y_{100} - Y_0$	P(Child in Q5| Parent in Q1)
1	Salt Lake City, UT	46.2	0.264	10.83%
2	Pittsburgh, PA	45.2	0.359	9.51%
3	San Jose, CA	44.7	0.235	12.93%
4	Boston, MA	44.6	0.322	10.49%
5	San Francisco, CA	44.4	0.250	12.15%
6	San Diego, CA	44.3	0.237	10.44%
7	Manchester, NH	44.2	0.296	10.02%
8	Minneapolis, MN	44.2	0.338	8.52%
9	Newark, NJ	44.1	0.350	10.24%
10	New York, NY	43.8	0.330	10.50%

Lowest Absolute Mobility In The 50 Largest CZs

Upward Mobility Rank	CZ Name	Y_{25}	$Y_{100} - Y_0$	P(Child in Q5| Parent in Q1)
41	Nashville, TN	38.2	0.357	5.73%
42	New Orleans, LA	38.2	0.397	5.12%
43	Cincinnati, OH	37.9	0.429	5.12%
44	Columbus, OH	37.7	0.406	4.91%
45	Jacksonville, FL	37.5	0.361	4.92%
46	Detroit, MI	37.3	0.358	5.46%
47	Indianapolis, IN	37.2	0.398	4.90%
48	Raleigh, NC	36.9	0.389	5.00%
49	Atlanta, GA	36.0	0.366	4.53%
50	Charlotte, NC	35.8	0.397	4.38%

Transmission intergénérationnelle de la pauvreté

Upward Mobility (Y_{25}) Adjusted for Differences in Cost of Living
Parent and Child Income Deflated by Cost of Living Based on ACCRA data



Corr. with baseline $\bar{y}_{25}$ = 0.98 (unweighted), 0.86 (pop-weighted)

Effets de la qualité de l'enseignement

Article : R. Chetty, J. Friedman et J. Rockoff, "Measuring the Impact of Teachers II : Teacher Value-Added and Student Outcomes in Adulthood", *American Economic Review* 104(9) : 2630-2679

Article : R. Chetty, J. Friedman et J. Rockoff, "Measuring the Impact of Teachers II : Teacher Value-Added and Student Outcomes in Adulthood", *American Economic Review* 104(9) : 2630-2679

Objectif :

- "Are teachers' impacts on students' test scores ("value-added") a good measure of their quality"

Effets de la qualité de l'enseignement

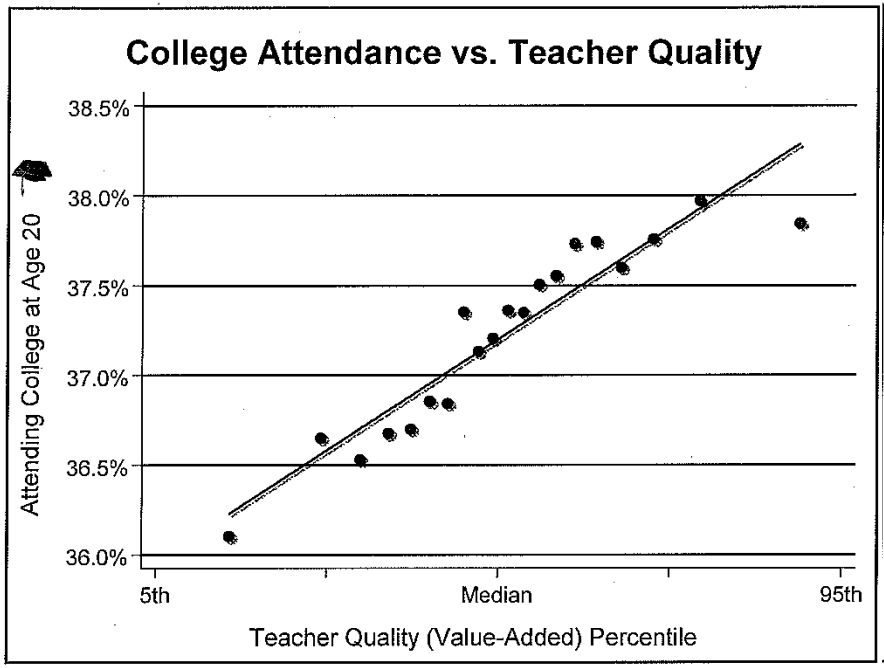
Objectif :

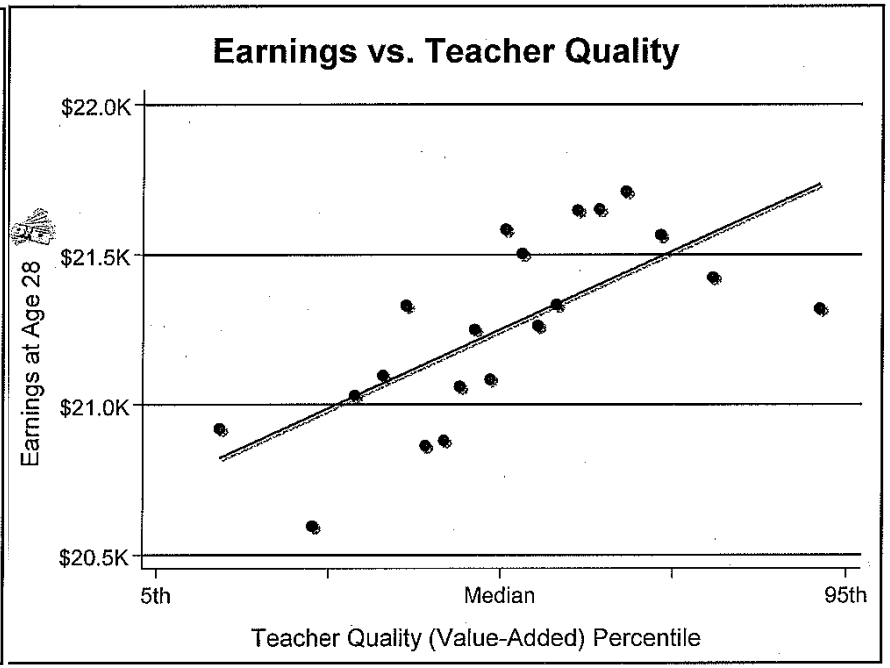
- "Are teachers' impacts on students' test scores ("value-added") a good measure of their quality"

Données :

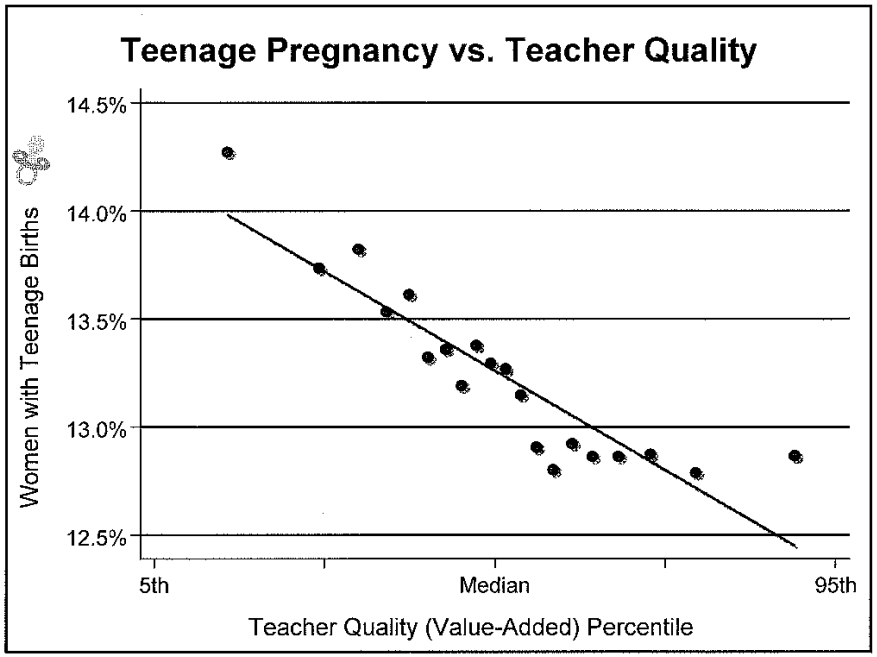
- Fusion de données administratives sur les étudiants et les professeurs en 3e–8e années pour les années 1980-2009 avec leurs données fiscales 1996-2011
- Information sur l'historique scolaire, résultats aux tests (Anglais, arts, math)
- 2,5 millions d'enfants, 18 millions de tests

Effets de la qualité de l'enseignement





Effets de la qualité de l'enseignement



Effets de la qualité de l'enseignement

Policy Impacts

"We know a good teacher can increase the lifetime income of a classroom by over $250,000.... Every person in this chamber can point to a teacher who changed the trajectory of their lives"

- Barack Obama, *State of the Union*, 2012

"A recent study by Harvard and Columbia economists found that students with effective teachers are less likely to become pregnant, more likely to go to college and more likely to get higher-paying jobs....Ineffective teachers are hurting our students' futures – we can't allow that."

- Michael Bloomberg, *State of the City*, 2012

Réforme de l'aide sociale en 1989 au Québec

Deux études :

1. Drolet, S., B. Fortin and G. Lacroix (2004), "Welfare Benefits, Minimum Wage Rate and the Duration of Welfare Spells : Evidence from a Natural Experiment in Canada", *Journal of Public Economics*, Volume 88, Issues 7-8, 1495–1520.
   $\Rightarrow$ Doubles différences
2. Lemieux, Thomas & Milligan, Kevin (2008), "Incentive effects of social assistance : A regression discontinuity approach", Journal of Econometrics, Elsevier, vol. 142(2), pages 807-828, February.
   $\Rightarrow$ Regression discontinuity

Réforme de l'aide sociale en 1989 au Québec

Deux études :

1. Drolet, S., B. Fortin and G. Lacroix (2004), "Welfare Benefits, Minimum Wage Rate and the Duration of Welfare Spells : Evidence from a Natural Experiment in Canada", *Journal of Public Economics*, Volume 88, Issues 7-8, 1495–1520.
   $\Rightarrow$ Doubles différences
2. Lemieux, Thomas & Milligan, Kevin (2008), "Incentive effects of social assistance : A regression discontinuity approach", Journal of Econometrics, Elsevier, vol. 142(2), pages 807-828, February.
   $\Rightarrow$ Regression discontinuity

Données

Fichiers administratifs du Ministère de la Solidarité sociale : 63 707 individus

Données

Fichiers administratifs du Ministère de la Solidarité sociale : 63 707 individus

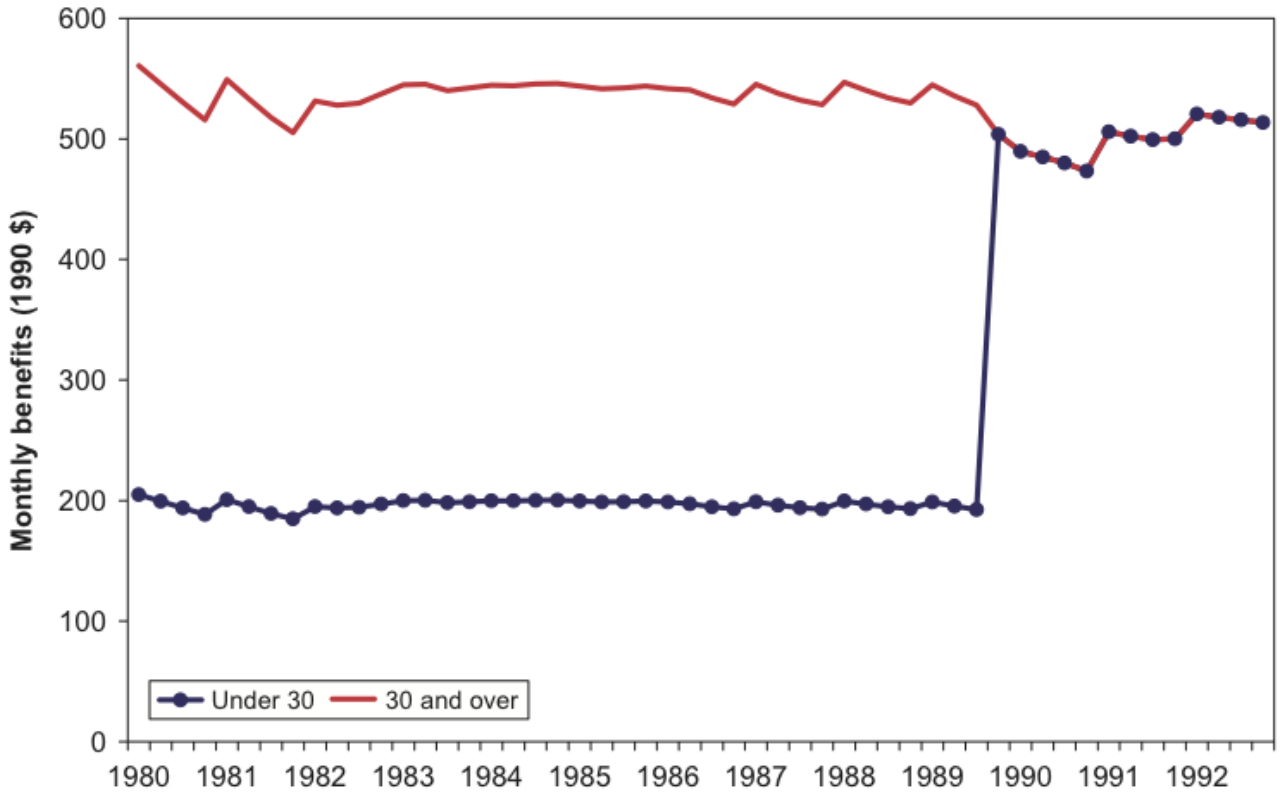
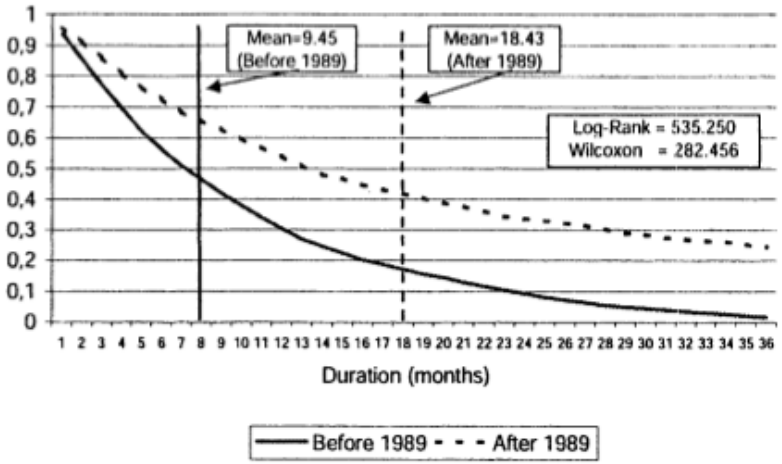


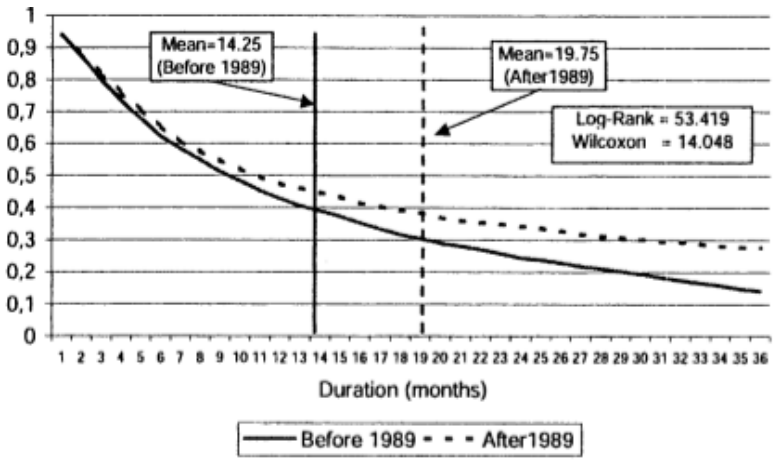
Fig. 1. Social assistance benefits, single individual.

Réforme de l'aide sociale en 1989 au Québec

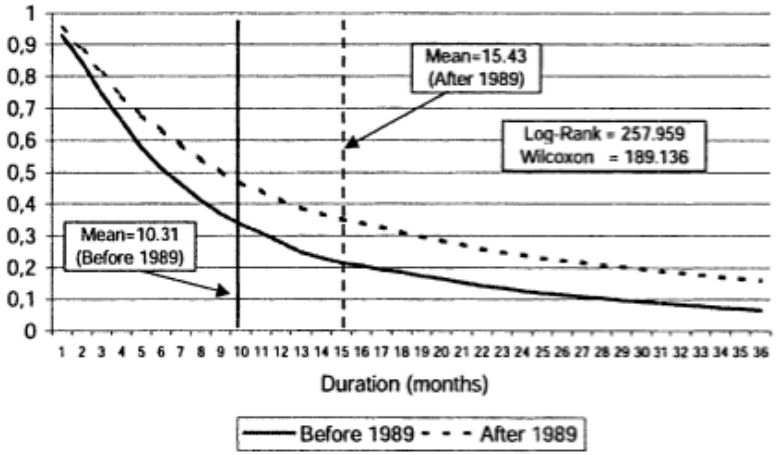
18–21 ans



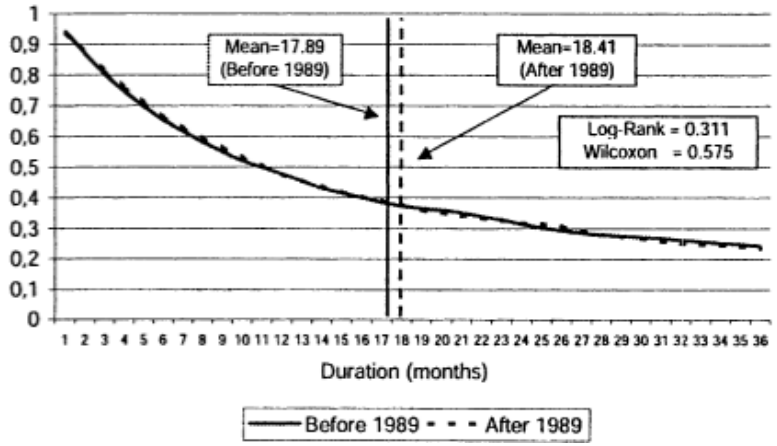
22–24 ans



25–29 ans



30 + ans



Article : Guy Lacroix et Thomas Vigneault, (2012) , “Subvention salariale et sortie de pauvreté”, Le Québec économique , Luc Godbout, Marcelin Joanis et Nathalie de Marcellis-Warin (eds)

Programme Action-Emploi

Article : Guy Lacroix et Thomas Vigneault, (2012) , “Subvention salariale et sortie de pauvreté”, Le Québec économique , Luc Godbout, Marcelin Joanis et Nathalie de Marcellis-Warin (eds)

Objectif :

Analyser l’impact du programme Action Emploi sur la dépendance à long terme des prestataires de l’assistance sociale

Le programme Action Emploi

- Permettre à des prestataires de longue durée de l'assistance-emploi de réintégrer le marché du travail de façon durable
- Entré en vigueur le 1er décembre 2001
- Avoir cumulé un minimum de 36 mois de présence à l'aide au cours des 45 mois précédant leur inscription à la mesure Action emploi
- Trouver un emploi 30 heures/semaine dans la première année
- Prestataires bénéficient d'un supplément à leurs revenus de travail pendant une période maximale de trois ans.
- L'aide accordée est modulée de manière à décroître avec le temps : 390 $/mois la 1ere année, 260 $/mois la 2e année, 130$/mois la 3e année

Données :

Fichiers administratifs du Ministère de la Solidarité sociale. Plus de 404 218 dossiers actifs en date du mois de décembre 2002, en plus de 13 308 participants au programme.

Le programme Action Emploi

Caractéristiques des prestataires de l'assistance sociale, selon le statut de participation à Action emploi	Participants à Action emploi (%)	Non participants (%)
Type de ménages		
Personnes seules	44,67	63,05
Familles monoparentales	29,82	13,88
Couples sans enfants	5,61	10,44
Couples avec enfants	19,71	12,45
Scolarité		
Primaire	5,60	15,70
Secondaire non complété	51,40	48,52
Secondaire complété	26,63	22,21
Postsecondaire	10,05	7,44
Université	6,21	6,13
Mois cumulatifs à l'assistance sociale		
De 36 à 47	5,28	1,76
De 48 à 119	31,71	9,96
120 et plus	59,41	80,99
Sexe		
Féminin	44,97	49,04
Masculin	55,03	50,96
Âge		
Moins de 25 ans	12,88	12,49
De 25 à 29 ans	10,98	7,13
De 30 à 34 ans	16,16	9,93
De 35 à 39 ans	18,49	12,34
De 40 à 44 ans	18,36	12,92
De 45 à 49 ans	13,04	12,30
De 50 à 54 ans	6,82	11,74
55 ans et plus	3,27	21,15

suite du tableau à la page sui

Caractéristiques des prestataires de l'assistance sociale, selon le statut de participation à Action emploi (suite)	Participants à Action emploi (%)	Non participants (%)
Région de résidence		
Bas-Saint-Laurent	4,00	3,06
Saguenay–Lac-Saint-Jean	3,63	4,13
Capitale-Nationale	11,58	8,22
Mauricie	6,79	4,93
Estrie	5,06	3,78
Outaouais	3,22	4,02
Abitibi-Témiscamingue	1,91	2,14
Côte-Nord	1,78	1,13
Nord-du-Québec	0,27	0,22
Gaspésie–Îles-de-la-Madeleine	3,21	2,02
Chaudière-Appalaches	3,43	3,44
Laval	2,67	2,67
Lanaudière	5,35	4,40
Laurentides	6,60	5,07
Montérégie	12,90	12,71
Centre-du-Québec	3,65	2,70
Montréal Banlieue	6,78	9,03
Montréal	17,13	25,12
Contraintes à l'emploi		
Sévères	4,26	32,94
Temporaires	22,58	28,19
Nombre d'observations	13 309	404 218

Tableau t/2012-c**10-1**

Le programme Action Emploi

Caractéristiques des prestataires de l'assistance sociale selon le statut d'admissibilité à Action emploi		
	Non admissibles (Cumul de 33 à 35 mois) (%)	Admissibles (Cumul de 36 à 38 mois) (%)
Type de ménages		
Personnes seules	59,43	58,01
Familles monoparentales	17,86	18,10
Couples sans enfants	7,86	8,81
Couples avec enfants	14,64	14,37
Scolarité		
Primaire	8,94	10,74
Secondaire non complété	50,74	50,02
Secondaire complété	23,51	23,60
Postsecondaire	9,69	8,97
Université	7,12	6,68
Mois cumulatifs à l'assistance sociale		
De 36 à 47	12,28	15,05
De 48 à 119	18,78	21,55
120 et plus	14,76	18,81
Sexe		
Féminin	50,71	50,78
Masculin	49,29	49,22
Âge		
Moins de 25 ans	18,38	16,86
De 25 à 29 ans	9,52	9,08
De 30 à 34 ans	12,32	11,48
De 35 à 39 ans	13,50	14,12
De 40 à 44 ans	13,69	13,43
De 45 à 49 ans	10,91	10,84
De 50 à 54 ans	8,98	9,44
55 ans et plus	12,69	14,75

Caractéristiques des prestataires de l'assistance sociale selon le statut d'admissibilité à Action emploi (suite)		
	Non admissibles (Cumul de 33 à 35 mois) (%)	Admissibles (Cumul de 36 à 38 mois) (%)
Région de résidence		
Bas-Saint-Laurent	2,60	2,70
Saguenay–Lac-Saint-Jean	3,80	4,26
Capitale-Nationale	8,68	8,53
Mauricie	4,58	4,67
Estrie	3,81	3,55
Outaouais	3,84	4,41
Abitibi-Témiscamingue	2,11	2,63
Côte-Nord	1,17	1,12
Nord-du-Québec	0,21	0,27
Gaspésie–Îles-de-la-Madeleine	1,58	1,86
Chaudière-Appalaches	2,92	3,15
Laval	2,92	2,72
Lanaudière	4,82	4,99
Laurentides	5,49	5,35
Montérégie	13,49	12,70
Centre-du-Québec	2,66	2,59
Montréal Banlieue	9,51	9,46
Montréal	24,65	23,81
Contraintes à l'emploi		
Sévères	9,86	12,29
Temporaires	20,22	22,18
Nombre d'observations	22 782	23 744

Le programme Action Emploi

	Effet net du programme Action emploi, en nombre supplémentaire de mois par année passés à l'assistance sociale										
	Années depuis l'implantation du programme										Taux de participation (%)
	1	2	3	4	5	6	7	8	9	10	
Groupe											
Échantillon complet	-0,22 *	-0,18 *	-0,12 *	-0,20 *	-0,23 *	-0,12 *	-0,08	-0,16 *	-0,05	-0,05	6,41
Sans contrainte à l'emploi	-0,36 *	-0,29 *	-0,18 *	-0,28 *	-0,35 *	-0,23 *	-0,16 *	-0,26 *	-0,17 *	-0,07	8,76
Nés au Canada	-0,41 *	-0,38 *	-0,32 *	-0,34 *	-0,40 *	-0,27 *	-0,20 *	-0,33 *	-0,23 *	-0,14 *	6,04
Nés hors Canada	-0,14	0,14	0,40 *	0,01	-0,13	-0,09	0,03	0,06	0,09	0,26 *	6,77
Région de résidence											
Montréal	-0,37 *	-0,16	-0,04	-0,24 *	-0,22 *	-0,27 *	-0,28 *	-0,48 *	-0,38 *	-0,22 *	5,84
Hors Montréal	-0,37 *	-0,41 *	-0,30 *	-0,32 *	-0,45 *	-0,21 *	-0,07	-0,10	-0,01	0,03	6,41
Type de ménage											
Personne seule	-0,42 *	-0,32 *	-0,17 *	-0,26 *	-0,36 *	-0,31 *	-0,25 *	-0,32 *	-0,20 *	-0,10	8,76
Monoparental	-0,52	-0,73 *	-1,26 *	-1,30 *	-1,22 *	-0,28	-0,25	-0,96 *	-0,60 *	0,05	11,08
Couple sans enfants	-1,27 *	-1,23 *	-0,38	-0,48	-0,58	-0,42	-0,11	-0,06	-0,60	0,34	6,42
Genre											
Femme	-0,30 *	-0,17 *	-0,07	-0,11	-0,16 *	-0,09	-0,06	-0,10	-0,01	-0,15 *	9,90
Homme	-0,44 *	-0,46 *	-0,34 *	-0,54 *	-0,62 *	-0,44 *	-0,30 *	-0,49 *	-0,40 *	0,05	8,01
Groupe d'âge											
Moins de 30 ans	-0,19	-0,14	-0,05	-0,76	-1,32 *	-0,96	-0,41	0,01	0,76	-1,16	7,64
De 30 à 44 ans	-0,43 *	-0,40 *	-0,26 *	-0,37 *	-0,49 *	-0,32 *	-0,23 *	-0,35 *	-0,35 *	-0,11	9,39
De 45 à 54 ans	-0,55 *	-0,61 *	-0,37	-0,39	-0,37	-0,29	-0,31	-0,47 *	-0,03	-0,01	4,20
Contraintes à l'emploi											
Contraintes temporaires	-0,08	-0,08	-0,14	-0,21 *	-0,20 *	-0,10	-0,12	-0,18 *	0,00	-0,07	4,66
Contraintes sévères	0,06	0,08	0,12	0,10	0,09	0,26	0,29	0,24	0,22	-0,06	1,59

Conclusions :

Conclusions :

Recherche

1. Les données massives présentent un potentiel énorme pour la recherche et l'évaluation des politiques publiques. Mais :
2. Le Québec accuse un sérieux retard sur le reste du Canada, les USA et l'Europe
3. La recherche fondée sur les enquêtes de moins en moins valorisée
4. Les chercheurs et les étudiants québécois sont pénalisés

Conclusions :

Recherche

1. Les données massives présentent un potentiel énorme pour la recherche et l'évaluation des politiques publiques. Mais :
2. Le Québec accuse un sérieux retard sur le reste du Canada, les USA et l'Europe
3. La recherche fondée sur les enquêtes de moins en moins valorisée
4. Les chercheurs et les étudiants québécois sont pénalisés

Formation

1. Repenser les curriculum académiques en science économique
   - "Machine Learning" : Pas uniquement réservé aux informaticiens/ingénieurs/statisticiens
   - Regression discontinuity, analyse non-paramétrique
   - Renforcement des connaissances en traitement informatique des grandes bases de données
2. Pertinente seulement si accès aux données, et si valorisée dans la