Essais d'évaluation de politique publique dans les champs de l'éducation, de la santé et des politiques d'emploi

Laurent Lequien

École des Hautes Études en Sciences Sociales

Thèse

Pour obtenir le grade de

Docteur de l'École des Hautes Études en Sciences Sociales

Discipline : Sciences Économiques

Présentée et soutenue publiquement le 9 février 2011 par

Laurent Lequien

Essais d'évaluation de politique publique dans les champs de l'éducation, de la santé et des politiques d'emploi

Directeur de thèse : M. Marc Gurgand

Composition du jury :

Rapporteurs :	M. Didier Blanchet	-	Administrateur de l'Insee
	Mme. Brigitte Dormont	-	Professeur à l'Université Paris Dauphine
Suffragants :	M. Marc Gurgand	-	Professeur à l'École d'Économie de Paris
	M. Edwin Leuven	-	Professeur à l'Ensae
	M. Thomas Piketty	-	Directeur d'Études à l'Ehess

Remerciements

Un grand merci à Marc Gurgand d'avoir accepté de diriger cette thèse, alors que le choix de mes thèmes de recherches était déjà bien arrêté. Ses conseils, toujours pertinents, sa capacité à analyser un sujet dans sa globalité comme dans ses moindres détails, m'ont été très précieux. Enfin, j'apprécie sa manière d'envisager son rôle d'économiste en prise directe avec l'actualité. Sa volonté constante d'apporter des éléments scientifiquement fondés pour éclairer la décision publique est un exemple pour moi.

Je tiens à remercier Didier Blanchet et Brigitte Dormont d'avoir accepté d'être rapporteurs de cette thèse, et Edwin Leuven et Thomas Piketty d'être membres du jury.

Des remerciements tout particuliers vont à Cédric Afsa, à l'origine de la plupart des études qui sont présentées dans cette thèse. Je le remercie de m'avoir proposé de travailler sur des domaines aussi variés que l'éducation, la santé et le marché du travail.

Un grand merci à Valérie Albouy, pour sa pédagogie et sa patience. Nous avons réalisé ensemble le chapitre 4 juste après mon arrivée à l'Insee, alors que je tentais pour la première fois d'appliquer les connaissances, principalement théoriques, acquises durant ma scolarité. Valérie m'a guidé autant que nous avons collaboré.

Je remercie également Bruno Crépon, Marc Gurgand et Thierry Kamionka de la confiance qu'ils m'ont témoignée en me proposant de faire partie de leur équipe pour réaliser une expérimentation aléatoire. Notre collaboration a été le fil rouge de cette thèse, puisqu'elle a commencé dès la fin de l'année 2005 et se poursuit encore aujourd'hui. Grâce à eux, j'ai découvert qu'un chercheur en économie ne travaille pas uniquement isolé dans son bureau, mais qu'il peut participer au quotidien à l'élaboration de politiques sociales innovantes.

Bien entendu, je remercie du fond du cœur toutes les personnes qui m'ont accompagné durant cette thèse à l'Insee, au Crest et à la Dares. J'ai été entouré de collègues ouverts et curieux, toujours prêts à s'intéresser aux problèmes que j'ai pu rencontrer. Merci pour ces innombrables conversations, aussi utiles qu'agréables. Un merci tout particulier à Sylvie Le Minez, qui m'accompagne dans mon parcours professionnel depuis plusieurs années.

Enfin, je tiens à remercier Florence d'être à mes côtés et de m'avoir donné une si belle raison de soutenir cette thèse en ce début d'année 2011.

Table des matières

Introduction générale

Quelle évaluation des politiques publiques ?

L'évaluation des politiques publiques a longtemps été absente du débat public en France, mais elle y occupe une place croissante depuis la fin des années 1990s. Cela est particulièrement vrai dans les domaines de l'emploi (Bonnal et al. 1997, Fougère et al. 2010) et de l'éducation (Piketty and Valdenaire 2006, Benabou et al. 2009). La France rattrape ainsi le retard pris sur certains pays anglo saxons (Etats Unis, Royaume Uni) ou d'Europe du Nord (Danemark, Pays-Bas, Suède, Finlande), où la culture de l'évaluation est très présente depuis plusieurs décennies (voir par exemple Woodbury and Spiegelman 1987, Gritz 1993, ou encore la création de la RAND corporation aux Etats Unis dès 1948).

Le mot "évaluation" peut généralement être interprété de plusieurs manières. Un commissaire-priseur évalue le prix d'un objet qu'il va mettre en vente ; un touriste au pied de la tour Eiffel évalue la hauteur de celle-ci. Un professeur évalue le niveau scolaire de ses élèves ; un démographe évalue le nombre de personnes habitant en France à une date donnée. Dans ces quatre exemples, l'évaluation consiste à déterminer la valeur d'une certaine quantité (un prix, une hauteur, un degré de compréhension, un effectif). Le point commun à tous ces exemples est que l'objet que l'on souhaite évaluer est directement observable, même s'il peut être difficile à mesurer en pratique. Il suffirait ainsi de compter le nombre d'habitants en France pour connaître sa population.

Dans le domaine des sciences économiques, beaucoup de paramètres d'intérêt sont également observables, grâce à la constitution de bases de données administratives. Les Déclarations Annuelles de Données Sociales (DADS) permettent par exemple de connaître le salaire de toutes les personnes travaillant dans le secteur privé en France. Le nombre de bénéficiaires du revenu de Solidarité active (rSa) est également accessible grâce aux données des caisses d'allocations familiales. Lorsque des sources administratives ne sont pas disponibles, des enquêtes peuvent être menées par le système statistique public et l'Insee en particulier.

Mais il arrive que le paramètre d'intérêt ne soit pas directement observable. C'est très souvent le cas lorsque l'on souhaite savoir si une politique publique est efficace. Pour illustrer cela, considérons la politique d'insertion dans l'emploi étudiée dans le chapitre 1. En 2005, un Conseil général a décidé de proposer un accompagnement personnalisé vers l'emploi aux allocataires du Revenu Minimum d'Insertion (RMI) de son département. La seule condition d'éligibilité était que ces personnes devaient percevoir ce minimum social depuis au moins deux années. Cet accompagnement a été réalisé par une société prestataire. Dans le cadre de son contrat, le prestataire s'est vu fixer un objectif de placement en emploi d'un certain nombre de personnes par an. Or, l'évaluation ne consiste pas à s'assurer que cet objectif a été atteint. Imaginons un cas extrême dans lequel toutes les personnes accompagnées par le prestataire trouvent un emploi à l'issue de la phase d'accompagnement, et où ces mêmes personnes auraient également trouvé un emploi en l'absence du programme d'accompagnement : dans ce cas (hypothétique), l'entreprise a strictement respecté ses engagements, pourtant son intervention n'a rien changé et la dépense engagée par le Conseil général est effectuée en pure perte. Imaginons au contraire qu'aucune des personnes placées par le prestataire n'aurait trouvé d'emploi sans accompagnement intensif : alors l'intervention a une forte valeur ajoutée, elle est très efficace et probablement très rentable pour la collectivité.

L'évaluation de cette politique d'emploi, dite évaluation d'impact, consiste à mesurer les valeurs prises par une variable d'intérêt en présence et en l'absence du dispositif, pour la même population, au même moment, et à les comparer. Autrement dit, elle consiste à mesurer ce que la mise en place du dispositif a changé, ou encore la valeur ajoutée du dispositif par rapport à une situation de référence. Ainsi, une évaluation d'impact est toujours une comparaison entre la situation qui prévaut en présence de la politique et la situation qui prévaudrait en son absence. Lorsque la variable d'intérêt est l'insertion dans l'emploi, comme dans l'exemple ci-dessus, il faut donc calculer le taux d'emploi des personnes prises en charge par le prestataire, et le comparer au taux d'emploi que ces mêmes personnes auraient eu si elles n'avaient pas bénéficié de la prise en charge. Cela mesure exactement l'impact causal que le programme d'accompagnement a eu sur le taux d'emploi des personnes qui ont bénéficié de l'accompagnement.

La notion de contrefactuel

La difficulté pour l'évaluateur est alors qu'il ne peut observer simultanément la situation d'une personne lorsqu'elle bénéficie de l'accompagnement et celle qu'elle aurait eu si elle n'en avait pas bénéficié. Il n'est alors pas possible de mesurer l'impact causal pour chacun des bénéficiaires, et d'en déduire ensuite un effet moyen sur l'ensemble des bénéficiaires. Estimer un impact causal nécessite donc de trouver un moyen d'approximer une quantité qui n'est pas accessible au statisticien : la situation qu'aurait connue le bénéficiaire du traitement s'il n'avait pas reçu ce traitement. Sans ce point de comparaison, la seule situation des bénéficiaires n'est pas informative de l'effet du traitement. Ce problème de *situation contrefactuelle* a été décrit par Rubin (1974) et apparait dans toutes les analyses causales, dépassant largement le cadre de l'évaluation des politiques publiques[1].

La manière la plus simple d'envisager une évaluation, au sens défini ci-dessus, pourrait consister à comparer la situation du groupe des personnes accompagnées avec la situation du groupe des personnes non-accompagnées. Cette démarche revient à supposer qu'en l'absence du programme d'accompagnement, les bénéficiaires du programme auraient connu la même situation que celle observée pour les non-bénéficiaires. Or cette hypothèse est ici très incorrecte. En effet, les personnes accompagnées sont, a priori, différentes de celles qui ne le sont pas. Ce sont celles que le prestataire est parvenu à contacter et qui ont accepté la proposition d'accompagnement : elles ont sans doute des caractéristiques peu semblables à celles des personnes qui n'ont pu être jointes ou celles qui ont refusé de se faire accompagner

[1] Rubin (1974) introduit ce concept avec un exemple très éloigné des politiques publiques. Il souligne la difficulté de mesurer l'impact causal de l'aspirine puisqu'on ne peut observer la même personne dire “Si j'avais pris il y a une heure deux aspirines au lieu d'un verre d'eau, mon mal de tête serait maintenant parti” et “Parce que j'ai pris il y a une heure deux aspirines au lieu d'un verre d'eau, mon mal de tête est parti”.

par le prestataire. Aussi, on ne sait pas, sans approfondir, dans quelle mesure l'intervention du prestataire a créé ces différences éventuelles de trajectoires, puisque les personnes accompagnées et non accompagnées sont a priori intrinsèquement différentes. Il n'est pas non plus d'un grand secours de comparer des bénéficiaires et des non-bénéficiaires de mêmes caractéristiques observables (âge, éducation, ancienneté dans le RMI, etc.), car ils diffèrent sans doute encore sous d'autres critères difficiles à observer et à objectiver (leur motivation à retrouver un emploi, par exemple). Si ces critères influent à la fois leur propension à accepter l'accompagnement et leur probabilité de trouver un emploi, ils créeront là aussi un différentiel d'emploi entre ces deux groupes qu'il ne sera pas possible de distinguer de celui dû à l'accompagnement.

Ce problème d'endogénéité de la participation au traitement se retrouve dans beaucoup de contextes. En plus des programmes dont la participation se fait sur la base du volontariat, la plupart de ceux dont l'éligibilité est basée sur un critère particulier créent naturellement une dissymétrie entre bénéficiaires et non-bénéficiaires : aide au logement ou bourse scolaire accessibles sous conditions de ressources, accompagnement vers l'emploi réservé aux jeunes, aux seniors, ou aux demandeurs d'emploi de longue durée, etc. Dans tous ces exemples, le simple fait de remplir le critère d'éligibilité peut en effet avoir un effet direct sur la variable d'intérêt. Prenons le cas d'une aide financière proposée aux ménages les plus pauvres pour acheter un logement et dont on souhaite étudier l'impact sur la probabilité d'accéder à la propriété. Comparer les personnes éligibles à l'aide à celles non éligibles n'est d'aucune utilité : la probabilité d'accession à la propriété étant par ailleurs corrélée avec le revenu des ménages, cette comparaison révèlera uniquement l'effet cumulé de l'aide et de cette corrélation, et ne permettra pas de mesurer séparément ces deux paramètres.

La méthode, séduisante car extrêmement simple, consistant à utiliser les non-bénéficiaires pour résoudre le problème de contrefactuel ne fonctionne donc généralement pas. Il faut alors mettre en œuvre des techniques plus sophistiquées pour identifier et mesurer l'impact causal d'un programme.

Plan de la thèse

La littérature propose plusieurs manières de réaliser une évaluation d'impact lorsque la participation est supposée endogène. Elles sont toutes basées sur le fait qu'il est nécessaire de disposer d'un facteur exogène expliquant (au moins partiellement) la participation au traitement pour obtenir l'identification. Cette source de variation exogène permet de corriger le biais qui pourrait exister entre les bénéficiaires et les non-bénéficiaires. Le choix d'une technique plutôt qu'une autre dépend à la fois du contexte, des sources d'identification et des données disponibles.

La suite de cette introduction présente les 5 évaluations étudiées dans cette thèse, ainsi que les 5 modélisations différentes mises en œuvre pour identifier l'impact causal. Le cadre économétrique permettant d'évaluer une politique publique est en effet différent dans chaque chapitre : expérimentation aléatoire (chapitre 1), doubles différences (chapitre 2), instrumentation sur données de panel avec un modèle structurel (chapitre 3), régressions par discontinuité (chapitre 4), et enfin le cadre classique des variables intrumentales (chapitre 5).

Les trois premiers chapitres portent sur le marché du travail, tandis que les deux derniers étudient l'existence de rendements de l'éducation sur la santé.

Accompagnement d'allocataires du Revenu Minimum d'Insertion

Le chapitre 1 présente l'évaluation par expérimentation aléatoire de la politique d'accompagnement vers l'emploi décrite plus haut, ainsi qu'une analyse coût/bénéfice de cette politique. Cette étude, réalisée en collaboration avec Bruno Crépon, Marc Gurgand et Thierry Kamionka, est à notre connaissance la première en France à mettre en œuvre un échantillonnage aléatoire pour évaluer l'impact d'une politique sociale.

Les expérimentations aléatoires s'inspirent des méthodes utilisées depuis longtemps dans le milieu médical. Elles sont basées sur une idée simple pour résoudre le problème de contrefactuel : face à des malades, c'est le hasard qui va décider qui recevra le médicament dont on souhaite tester l'efficacité. En donnant par exemple une chance sur deux à chaque malade de recevoir le nouveau médicament, on crée ainsi deux groupes de patients : le premier groupe va recevoir le traitement à tester, alors que l'autre groupe va recevoir le médicament habituel. L'affectation aléatoire des individus dans l'un des deux groupes est l'instrument qui permet d'identifier l'impact causal. Elle crée en effet deux groupes de personnes qui sont a priori parfaitement identiques. La seule dimension qui les différencie est que l'un des groupes va bénéficier du médicament innovant, tandis que l'autre reçoit le médicament de référence. Il suffit alors de comparer le taux de guérison dans les deux groupes pour savoir quel médicament est le plus efficace. Recevoir ou non le traitement à évaluer n'est plus endogène grâce à l'introduction d'une affectation aléatoire dans l'un des deux groupes.

La première utilisation de cette comparaison entre un groupe de traitement et un groupe de contrôle est attribuée à un médecin écossais, James Lind (Lind 1753, Duncan 1997). En 1747, celui-ci a ainsi mis en évidence l'effet bénéfique d'un régime à base d'oranges et de citrons pour soigner des marins atteints de scorbut. Il faut ensuite attendre que Fisher (1935) pose les bases théoriques des expérimentations aléatoires dans un livre qui fait encore aujourd'hui référence. Fisher (1935) définit précisément ce qu'est une expérimentation aléatoire, le jeu d'hypothèses nécessaires pour qu'elle soit valide, et à quel type de questions une expérimentation donnée peut a priori répondre. Il y discute également l'extrapolation que l'on peut faire des résultats (réplication, etc.), et l'efficacité d'une telle méthode. Aujourd'hui, les expérimentations aléatoires occupent une place de plus en plus importante en sciences sociales (voir par exemple Banerjee and Duflo 2008).

Le grand avantage des expérimentations aléatoires est leur apparente simplicité : il suffit de répartir aléatoirement les personnes en deux groupes, de donner le traitement à l'un des groupes, et ensuite d'observer si le comportement des deux groupes diffère pour savoir si le traitement a un impact significatif ou non. Une limite claire est cependant qu'il n'est pas toujours possible de procéder à un tirage au sort pour déterminer qui recevra le traitement. Soit que des raisons éthiques l'interdisent, soit que l'organisation pratique de ce tirage soit rendue impossible par un contexte particulier[2].

[2]Voir Banerjee and Duflo (2008) et Deaton (2008) pour une discussion détaillée sur les avantages

En accord avec le Conseil général, nous avons d'abord constitué aléatoirement un groupe de traitement et un groupe de contrôle au sein de l'ensemble des allocataires du RMI éligibles à ce dispositif. Le prestataire chargé de l'accompagnement devait contacter les personnes du groupe de traitement pour leur proposer cet accompagnement personnalisé. Il ne pouvait contacter le groupe de contrôle car il ne disposait pas de leurs coordonnées.

En théorie, chaque personne du groupe de traitement était donc accompagnée par le prestataire, et aucune du groupe de contrôle ne l'était. En pratique, certaines personnes assignées au traitement ont refusé de suivre le traitement, et inversement des personnes du groupe de contrôle ont demandé expressément à être accompagnées alors qu'elles ne devaient théoriquement pas l'être. Le fait que l'on puisse évaluer la prestation par expérimentation aléatoire, même si la règle du tirage au sort n'est pas strictement respectée (*imperfect compliance*), a donc été crucial d'un point de vue opérationnel. Il est en effet difficile de mettre en œuvre des protocoles qui imposent strictement une situation aux individus : il n'est ni éthiquement convenable, ni techniquement faisable, d'imposer (ou de refuser) un accompagnement à une personne.

Le protocole d'évaluation par expérimentaion aléatoire permet cependant de tenir explicitement compte d'un tel cas de figure, même si cela a pour effet de réduire la précision statistique et de compliquer légèrement le raisonnement. Au lieu d'estimer l'effet moyen du traitement sur l'ensemble des traités, on identifie alors un effet moyen local (*local average treatment effect*, LATE, d'après la terminologie proposée par Angrist and Imbens 1994) sur un sous-ensemble des traités (les *compliers*, toujours d'après Angrist and Imbens 1994) : les personnes qui acceptent de recevoir le traitement lorsqu'on les affecte au groupe de traitement, et qui ne demandent pas à le recevoir lorsqu'on les affecte au groupe de contrôle.

Une particularité de l'expérimentation présentée au chapitre 1 est qu'un certain laps de temps peut s'écouler entre le tirage au sort et l'entrée en traitement. Dans un tel cadre d'entrée dynamique, la notion de *complier* devient fluctuante, car une même personne peut être *complier* à une date donnée, et *non-complier* à la

et inconvénients des expérimentations aléatoires.

date suivante. Cet aspect dynamique complique l'identification de l'impact causal par rapport au cadre général présenté par Angrist and Imbens (1994). Nous montrons que le LATE peut être estimé lorsque l'impact causal dépend de la durée en traitement, ou lorsque l'impact causal est corrélé avec la probabilité d'entrée en traitement, mais que le modèle n'est pas identifié lorsque ces deux hypothèses sont posées simultanément.

L'impact de la durée d'un arrêt de carrière sur les salaires

Les chapitres 2 et 3 sont consacrés à la trajectoire salariale des femmes en France, et plus précisément des femmes qui ont eu des enfants une fois entrées sur le marché du travail. Après avoir donné naissance à un enfant, beaucoup de femmes se retirent du marché du travail plus longtemps que la durée légale du congé maternité. Ces deux chapitres mettent en évidence que la durée de ce retrait temporaire semble avoir un impact négatif sur le salaire perçu après le retour à l'emploi, et quantifient l'ordre de grandeur de cet effet.

L'identification de cet effet causal repose sur la réforme de l'Allocation Parentale d'Education (APE) qui a eu lieu en 1994. L'APE est une somme forfaitaire versée aux jeunes parents qui ne travaillent pas à temps plein, et ce jusqu'au troisième anniversaire de l'enfant. Jusqu'en juin 1994, l'APE n'était accessible que pour une naissance de rang 3 ou plus. Les parents d'un deuxième enfant ne bénéficiaient donc pas d'incitation particulière pour réduire temporairement leur offre de travail après la naissance. A partir de juillet 1994, l'APE leur permet de compenser partiellement la perte de revenu occasionnée par un retrait du marché du travail.

La réforme de l'APE permet l'identification de l'impact causal à condition que deux critères soient satisfaits. D'une part, cette réforme doit avoir modifié le choix de la durée de l'arrêt pour un grand nombre de salariées. Piketty (2005) et Pailhé and Solaz (2006) ont observé que cette réforme a incité un grand nombre de femmes à diminuer leur offre de travail dans les trois années suivant la naissance de leur deuxième enfant. La partie 2.1.2 confirme ces résultats et montre que l'extension de l'APE en 1994 a effectivement modifié la distribution des durées d'interruption de carrière. D'autre part, elle ne doit pas avoir eu d'influence sur les autres variables individuelles ayant un impact direct sur le salaire (être plus ou moins fortement carriériste, avoir de bonnes perspectives de salaire, etc.). Autrement dit, l'unique manière pour ce changement législatif d'affecter les salaires est d'activer un mécanisme causal indirect passant par la durée de l'arrêt : il modifie la durée de l'arrêt, et ce changement de durée a à son tour un impact sur le salaire.

Ainsi, si l'on constate une différence de salaire entre les femmes selon qu'elles ont eu leur enfant sous l'ancien ou le nouveau régime législatif, ce différentiel pourra être attribué à la variation de la durée du retrait du marché du travail. En revanche, si le niveau de salaire ne change pas avec l'application de la réforme alors que la durée a été modifiée, cela signifie forcément que la durée de l'arrêt n'a pas d'effet sur les salaires. Cette méthode permet donc de savoir si cet impact causal existe réellement, et dans l'affirmative, de le quantifier.

Le chapitre 2 met en œuvre une méthode de doubles différences pour réaliser l'évaluation d'impact. Cette stratégie consiste à comparer la trajectoire salariale des femmes qui ont eu un deuxième enfant juste après la réforme à celle des femmes qui ont eu leur deuxième enfant juste avant la réforme. Ces deux groupes de femmes sont censés être en tous points similaires, si ce n'est que le premier a bénéficié de l'APE et que le second n'en a pas bénéficié. Les doubles différences suggèrent qu'une durée d'interruption plus longue diminue le salaire lors du retour à l'emploi. Ce résultat n'est cependant valable qu'à condition que les conditions macroéconomiques (qui pourraient influencer le salaire, ou la durée d'interruption via le chômage) aient été les mêmes pour ces deux groupes de femmes. Il est possible de lever cette dernière hypothèse en utilisant un groupe de contrôle supplémentaire, par exemple les femmes ayant eu un premier enfant juste avant ou juste après la réforme de l'APE. Une estimation en triples différences donne alors des résultats qualitativement similaires à ceux obtenus en doubles différences, bien que moins précis.

Le chapitre 3 étudie la même question qu'au chapitre 2 : la durée d'un arrêt de carrière après une naissance a-t-elle un impact sur le salaire des mères lorsqu'elles reviennent sur le marché du travail ? La manière d'aborder ce problème y est cependant radicalement différente. Au lieu d'une approche non paramétrique par doubles ou triples différences, ce chapitre repose sur une modélisation beaucoup plus structurelle, proposée par Semykina and Wooldridge (2010). Exploitant la dimension longitudinale des données, nous décrivons la décision de participer au marché du travail année après année, et estimons ensuite une équation de salaire pour les femmes qui travaillent une année donnée. Il s'agit donc d'une estimation à la Heckman (1976) pour corriger du biais de sélection dans l'équation de salaire créé par la décision de participer au marché du travail. Deux particularités rendent ce modèle original : les données de panel permettent d'estimer un effet fixe individuel pouvant être corrélé avec les autres explicatives à la fois dans l'équation de sélection et dans l'équation de salaire, et cette dernière contient une variable endogène (la durée d'interruption). Ce problème d'endogénéité est résolu en instrumentant par le fait que la deuxième naissance ait eu lieu avant ou après la réforme de l'APE en juillet 1994.

Les résultats produits par l'estimation de ce modèle économétrique sont similaires à ceux obtenus en doubles différences : l'impact causal sur les salaires est négatif. Ce paramètre est du même ordre de grandeur que dans le chapitre 2, et est maintenant estimé avec une meilleure précision. Il semble donc qu'une durée d'interruption plus longue cause une baisse de salaire après le retour sur le marché du travail. A notre connaissance, ce résultat n'avait pas encore été montré sur données françaises.

De plus, il est rassurant de constater que les deux méthodes utilisées aux chapitres 2 et 3 conduisent aux même résultats, alors que les hypothèses sous-jacentes à chacun des modèles diffèrent. Les paragraphes 2.2.1 et 3.6 décrivent ces hypothèses et discutent les avantages et inconvénients respectifs des modèles structuels et non structurels.

Les rendements de l'éducation sur la santé

La dernière partie de cette thèse est consacrée au lien qui pourrait exister entre le niveau d'éducation d'une personne et son état de santé. Plus précisément, elle tente de déterminer si un plus haut niveau d'éducation, délivré à une même personne, produit dans le cours ultérieur de sa vie un effet protecteur sur sa santé. La mesure de la santé utilisée dans cette partie est la survie à un âge donné.

L'existence d'un lien empirique entre éducation et santé est unanimement constatée dans la littérature (Grossman 2004). Bien qu'il ne se dégage pas de consensus clair, une majorité d'auteurs considère qu'un mécanisme causal explique la corrélation observée dans les données[3]. La démonstration empirique de cette causalité ne va cependant pas de soi. La difficulté de la comparaison entre individus aux niveaux d'éducation différents est qu'on ne sait pas ce qui les a conduit à choisir leur niveau d'éducation. Il peut en effet y avoir des caractéristiques qui expliquent à la fois leur niveau d'éducation et leur état de santé, telles que les préférences pour le présent ou l'environnement familial (Grossman 1972, Fuchs 1982).

Le chapitre 4, réalisé en collaboration avec Valérie Albouy, estime des rendements de l'éducation sur la santé dans un cadre de régressions par discontinuités. La première exploitation de discontinuités dans les probabilités de recevoir un traitement est assez ancienne : Thistlethwaite and Campbell (1960) évaluent l'impact des bourses sur les aspirations en terme de carrière des étudiants. Ils utilisent le fait qu'aux États-Unis les bourses au mérite sont attribuées en fonction des résultats scolaires mais que cette relation n'est pas continue : le montant des bourses octroyées et les résultats scolaires sont reliés par une fonction en escalier. Leur intuition est que les étudiants ayant des résultats scolaires juste en dessous d'un saut de la fonction en escalier sont comparables à ceux situés juste au-dessus du seuil, seul le montant de bourse perçu les différenciant. Donc, s'ils constatent une différence dans les aspirations en termes de carrière entre les étudiants juste au-dessus du seuil et ceux juste au-dessous, cette différence pourra être attribuée au montant de la bourse car c'est la seule caractéristique qui distingue ces deux groupes de personnes.

Plus récemment, cette méthode a été utilisée pour évaluer l'impact de la taille des classes sur les performances scolaires des élèves (Angrist and Lavy 1999) ou estimer la sensibilité des étudiants aux incitations financières dans leur choix pour rejoindre une université américaine donnée (van der Klaauw 2002). Des discontinuités existent

[3]Se reporter au chapitre 4 pour une revue de la littérature.

notamment lorsque l'éligibilité au traitement est fonction d'un barème ou cible une partie de la population : l'aide sociale n'est octroyée que sous un seuil de revenu, l'aide à l'embauche des seniors qu'à partir d'un certain âge. La discontinuité dans la probabilité de recevoir le traitement peut également être temporelle et venir d'un changement dans les règles d'éligibilité à ce dernier ; ainsi, une approche classique d'évaluation d'une mesure de politique publique repose sur la comparaison de la situation juste avant et juste après la mise en place de la mesure.

Si les applications reposant sur cette idée de discontinuité dans le traitement se sont largement multipliées ces dernières années, l'article définissant proprement le cadre théorique est relativement récent (Hahn et al. 2001). Ces derniers précisent notamment quelles sont les conditions minimales d'identification d'un effet causal, et proposent une approche non paramétrique pour estimer les effets.

Le chapitre 4 exploite les discontinuités sur le niveau d'éducation créées par deux réformes du système éducatif français, qui modifient successivement l'âge de fin de scolarité obligatoire. La première réforme a porté cette obligation scolaire à 14 ans révolus pour les personnes nées en 1923 et après, alors que les individus nés avant cette date pouvaient quitter l'école dès 13 ans. En 1959, cette durée de scolarisation obligatoire a de nouveau été allongée, de deux ans cette fois-ci, pour les personnes nées après le 1^{er} janvier 1953. Ce type de changement législatif a été utilisé dans de nombreux autres pays pour identifier des rendements de l'éducation (voir par exemple Lleras-Muney 2005, Oreopoulos 2006).

La comparaison des taux de survie des générations nées de part et d'autre de la réforme ne met pas en évidence d'impact causal significatif de l'éducation sur la santé, que la survie soit mesurée à 50 ans ou à 80 ans. Cela indiquerait que la santé des compliers de ces deux réformes ne s'est pas améliorée parce que la loi les a obligés à aller à l'école plus longtemps qu'il ne l'auraient souhaité. Ce résultat s'insère dans une littérature indécise quand ce type d'instrument est utilisé sur données étrangères : certaines études concluent à l'absence d'impact causal (Arendt 2005, Doyle et al. 2005) tandis que d'autres trouvent un impact positif et significatif sur la santé (Glied and Lleras-Muney 2003, Oreopoulos 2006).

Le chapitre 5 utilise une variable instrumentale totalement différente pour estimer des rendement de l'éducation. Cette source d'identification, plus originale qu'une réforme sur l'âge de fin de scolarité obligatoire, repose sur les désordres créés en France par la seconde guerre mondiale. La France a en effet été fortement impliquée dans la seconde guerre mondiale, et la société française dans son ensemble a subi une profonde désorganisation pendant plusieurs années. Ce chapitre commence par montrer que les personnes qui étaient adolescentes pendant la guerre ont quitté l'école plus tôt qu'elles ne l'auraient fait en l'absence de la guerre. Ces générations les plus touchées par la guerre dans leurs choix éducatifs constituent donc notre groupe de traitement sur lequel l'instrument va jouer.

L'hypothèse d'exclusion prend ici une forme particulière. Il faut en général que l'instrument n'ait pas d'impact direct sur la santé du groupe de contrôle. Ici cette hypothèse n'est clairement pas vérifiée, puisqu'il semble évident que les privations et le stress engendrés par la guerre ont directement affecté la santé de toutes les personnes vivant à cette époque. Une hypothèse plus faible permet cependant l'identification : il suffit que l'impact direct de la guerre sur la santé soit le même pour toutes les personnes de l'échantillon. L'appartenance au groupe de traitement étant définie par l'année de naissance, cela revient à supposer que l'effet que la guerre a eu sur la santé (excepté via son impact sur le niveau d'éducation) n'a pas dépendu de l'âge des individus durant le conflit[4]. La section 5.5 propose une discussion détaillée de la validité de cette hypothèse, et conclut sur sa probable validité.

Les estimations par variable instrumentale indiquent que l'éducation a un impact causal significatif sur l'état de santé : une année d'études supplémentaire augmente la probabilité d'être encore en vie à 60 ans de 6 points de pourcentage, pour les personnes affectées par l'instrument. L'ampleur de ce coefficient est dans la fourchette haute des résultats présents dans la littérature, parmi les articles qui mettent en évidence un impact causal.

Le chapitre 5 semble à première vue apporter une réponse en contradiction avec celle du chapitre précédent. Elles devraient pourtant être comparables : ces cha-

[4]Cet échantillon est composé de personnes nées entre 1920 et 1945.

pitres portent sur le même pays, exploitent la même base de données, et utilisent la même mesure de la santé (la survie à un âge donné). Seuls les instruments diffèrent. Ce dernier point a de l'importance, car d'après Angrist and Imbens (1994), les estimations produisent un LATE identifié sur les compliers de chacun des instruments. L'interprétation de ces deux résultats est donc forcément très différente, puisqu'ils ne sont pas estimés sur les mêmes compliers. Les compliers du chapitre 4 sont ceux qu'une réforme a forcé à prolonger leur scolarité plus longtemps que souhaité. Dans le chapitre 5, les compliers ont au contraire dû quitter l'école avant d'avoir atteint ce qui aurait été leur niveau scolaire optimal en temps de paix. Une dimension qui les différencie clairement est donc leur motivation à continuer les études dans leurs dernières années de scolarité. Cette constatation suggère que les rendements de l'éducation sur la santé pourraient dépendre de la motivation des élèves à acquérir des connaissances et du capital humain à l'école. Cette possible hétérogénéité des rendements de l'éducation n'a à ma connaissance pas encore été explorée.

La conclusion rappelle les principaux résultats de cette thèse, et discute l'interprétation et la portée de ces travaux. Elle évoque enfin leurs limites, qui ouvrent de nouvelles perspectives de recherche.

Première partie

Marché du travail

Chapter 1

Is Counseling Welfare Recipients Cost Effective ? Lessons from a Random Experiment

---

Sommaire

This chapter was written with Bruno Crépon (Crest-Insee and JPAL), Marc Gurgand (Paris School of Economics, Crest-Insee and JPAL) and Thierry Kamionka (Crest-Insee).

1.1 Introduction

Active labor market policies have received increasing attention over the last decades, both from policy makers and researchers. Training programs have been extensively studied[1], and so have incentive policies[2]. More recently, policy interest has shifted to monitoring and counseling job-search, because it tends to directly decrease market frictions (rather than indirectly through financial incentives), an approach that has some socially desirable properties: Boone et al. (2007) argue that decreasing unemployment benefits is a second best policy when effort is not observable, but is dominated by monitoring and counseling if job-search can be observed and influenced *at reasonable cost.* The extent to which employment services are able to modify people's search effort or efficiency is an empirical issue. As indicated by Card et al. (2010), evaluation of such policies is rather positive. For instance, Crépon et al. (2005), Behaghel et al. (2009), Fougère et al. (2010) find that counseling policies implemented in France increase job finding; Dolton and O'Neill (1996, 2002) and Blundell et al. (2004) find similar results in Great Britain. The main exception is van den Berg and van der Klaauw (2006), who find no effect of counseling and monitoring based on a random experiment in the Netherlands. However, even when they do affect transition to work, those policies can be extremely costly, because they usually require hours of individual contact with a counselor or a case worker, and one may wonder if they are sustainable. Cost-effectiveness has never been formally evaluated in this literature, which therefore lacks a metric to judge whether effects are high or low.

This paper considers a job-search counseling program targeted at individuals that have been on welfare for at least two years, in a French district. Since January

[1]Well known evaluations include Gritz (1993), Bonnal et al. (1997), Gerfin and Lechner (2002) or Sianesi (2004).

[2]See, for instance, Woodbury and Spiegelman (1987), Decker and O'Leary (1994), Blundell (2006), and the aboundant literature on the Canadian Self-Sufficiency project: Card et al. (2005), Card and Robins (2005), Card and Hyslop (2005), Kamionka and Lacroix (2008), Brouillette and Lacroix (2010).

2004, French districts ("*Départements*") are in charge of welfare recipients and can propose innovative policies. In 2005, this district selected a private provider to run the counseling program. Access to the labor market is particularly difficult for the target population: it is generally very poorly educated, often suffers from healthcare problems, has long unemployment spells, has lost contact with the labor market, etc. Whether intensive counseling can be helpful to this particular population is an open question. We evaluate the impact of this policy on job access, using a random allocation of welfare recipients to the program. We also evaluate its cost-effectiveness. An argument for granting the program to a private provider, rather than rely on existing public social services, is that its increased effectiveness will compensate the additional cost, thanks to the induced reduction in welfare payment. We thus compare the financial net flows in the treated and control populations. Furthermore, by assigning reasonable social values to the employment status of former social recipients, we estimate a social net benefit of the policy.

We find some impact of this policy on employment and welfare transfers, and stronger ones among the population that has the shortest seniority on welfare. However, the cost of the policy is so high that it is never compensated by welfare payment reduction. To give orders of magnitude, welfare recipients receive on average about 1,000 € per trimester, whereas the provider receives a lump-sum payment of 2,200 € per enrolment into the program, plus additional payment per trimester when the person remains in employment. Even in the absence of additional payment, break even would require a causal impact of more than 2 trimesters of employment, which is far from what we estimate. Overall, we typically find the net cost to be above 2,000 €. Even when a social value of employment is included, the policy fails to have a positive return. The policy is more efficient on individuals with "only" 2 to 4 years of seniority in welfare, but the net cost remains high at about 1,700 €.

In the French context, this finding remains hard to interpret, because the counseling market has only recently been opened to the private sector: it is possible that private providers price over their marginal cost in a market in infantry where competition is limited. On another hand, it is also possible that private providers are not particularly efficient: based on another random experiment, Behaghel et al.

(2009) have shown that the public service is both more efficient and less costly at providing counseling to the unemployed. Generally, this result emphasizes that counseling and monitoring are very intensive policies, for which cost-effectiveness is an important issue.

On the methodological side, we discuss identification of a treatment effect when entry into the treatment can occur some time after randomization has taken place. In our design, a list of individuals is assigned into either control or treatment group. The treatment list is handed to the provider of counseling, but it takes time to get in touch with welfare recipients. Furthermore, some individuals from the control group may ask for treatment, and cannot be refused. Compliance to the random assignment is not perfect, and we must use some instrumental variable estimation. It is well known that, when there is essential heterogeneity[3], the estimated parameter is a Local Average Treatment Effect (Angrist and Imbens 1994), and must be interpreted as the causal impact on a population of compliers. However, we show that, when there is dynamic entry into treatment, the LATE interpretation becomes more complicated. The reason is that a person from the treatment group can be a never taker at some date (as long as she did not enter treatment) and a complier later on (once she has entered treatment). Similarly, a person from the control group can be a complier, then an always taker. Therefore, the estimated impact parameter can be different at every point in time, because it is identified on a different population of compliers.

As a consequence, a model where the structural policy impact is changing with time since entry into treatment cannot be identified under essential heterogeneity. The reason is that both essentiel heterogeneity and such time-dependance of the impact imply parameter changes over time. One has to choose to assume either essential heterogeneity or time-varying impact. Depending on the assumption, parameter interpretation and estimation methods both differ. This is a different issue from the ones analyzed by Abbring and van den Berg (2005).

We first present the policy, the experimental design and the data. We then

[3]i.e. when treatment impact is heterogenous and correlated with the propensity to enter treatment.

discuss identification and estimation issues along the lines just summarized. Results on employment status and cost-effectiveness are then presented. The final section concludes.

1.2 Policy and experimental design

1.2.1 A counseling scheme

Since 2004, French districts ("*Départements*") manage the welfare system that consists mainly in the payment of a guaranteed minimum income ("*Revenu Minimum d'Insertion*") to individuals with no or very low income, and in providing counsel and support to welfare recipients, traditionally either through in-house services or with the support of NGOs. As soon as 2005, an urban district decided to spend substantial resources on employment policy for welfare recipients living in that district. Its goal was to provide help and counsel to those likely to have the most severe difficulties on the job market and, it believed, had few chances to obtain a job without intensive assistance. The policy was thus targeted at those with at least 2 years of seniority into the welfare system. These recipients could be registered at the National Employment Agency and actively looking for a job, but had no obligation to do so.

As a novel feature at the time, the district elected assembly decided to select a private operator to implement this intensive counseling. The private market for job-search counseling was just starting to develop in France, particularly under the initiative of the Unemployment Benefit Agency (*Unédic*). Few operators had yet invested the area, and few were in a position to answer the tender.

The operator was supposed to contact a list of eligible individuals (on welfare since at least two years), and counsel them until they had found a job. With a letter or a phone call, the company explained that it contacted them on behalf of the district authority to help them find a job. It also invited them to a collective meeting where they would have more information on the counseling scheme and could enlist into the programme. Enrollment was voluntary, and welfare recipients could refuse to attend this meeting or enter treatment. When individuals were enrolled, the provider would allocate them a unique counselor that would meet him or her at least once a week in the company's offices. Once the person was employed, the counselor would follow up on the job.

Payment to the operator included a lump sum for each individual counseled, and increasing bonuses if she worked during 3, 6 or 9 consecutive months between her enlistment and the end of the experiment in October 2007. The lump sum was 2,200 €; additional payment was 1,600 € after 3 months employment, plus an additional 500 € after 6 months and 500 € more after 9 months.

1.2.2 Experimental design and take-up

From the outset, the district considered that this policy was experimental. The decision to continue or not such a policy was conditioned by the results obtained during the experimental period. Clearly, the district was unsure about the effectiveness of the policy. In order to provide rigorous evaluation, we proposed to assign individuals randomly into the list that would be transmitted to the operator. Randomization took place between March and October 2006. All welfare recipients in the district with at least two years of seniority into the welfare system at the moment of randomization were eligible. Each eligible individual had a 75% chance of being affected to the treatment group, and 25% chance to the control group.

The first random draw occurred in March 2006, and concerned the 14,980 welfare recipients fulfilling the seniority requirement between December 2005 and February 2006. Among them, 11,222 individuals were randomly assigned to the treatment group, and the rest (3,758) to the control group. This constituted the first wave of entry into the experiment.

The second wave included welfare recipients who became eligible between March and May 2006: 1,176 individuals were randomly affected to the treatment group, and 395 to the control group. Likewise, a third random draw concerned those becoming eligible between the second wave and October 2006: 838 persons were added to the treatment group, and 277 to the control group. From November 2006 on, entry into the experiment was not possible anymore.

These three waves of entry cumulate a total of 17,666 individuals taking part into the experiment. The treatment group includes 13,236 individuals, and 4,430 belong to the control group. As expected, this repartition is extremely close to the theoretical proportions induced by a 75% / 25% random draw.

The experiment ended in October 2007. Therefore wave 1 was present in the experiment during 6 trimesters, whereas waves 2 and 3 lasted only 5 and 4 trimesters respectively.

All our data are from administrative files on welfare recipients. They include a limited set of individual characteristics, but an accurate series of transfer payments, and an employment status variable which will be discussed later. This information is complemented with the treatment/control variable that we generated, and it is merged with the list of individuals that entered treatment at every point in time. This list is established by the provider for billing purposes, therefore it is very reliable.

Table 1.1 presents descriptive statistics on the two randomized groups: proportion of women, age, number of children, seniority into the welfare system, employment rates in the last two months before entry into the experiment, and the amount of welfare transfer received during the four trimesters before entry into the experiment. As expected, these two groups are statistically similar along all dimensions. One noticeable fact is that the first wave constitutes 85% of the total sample.

Figure 1.1 plots the cumulative distribution of seniority in wave 1, measured at entry into the experiment in March 2006. Half of this subsample has between 2 and 4 years of seniority in the welfare system. 35% have more than 6 years of seniority, and some individuals are present in the welfare system since its creation. As will be observed later, this is a source of significant heterogeneity in treatment impact.

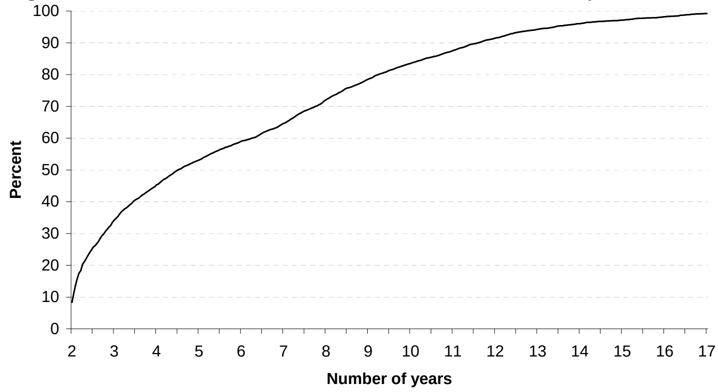
The private operator was to offer counseling to all individuals in the treatment group. In reality, enrollment was on a voluntary basis, and a large majority of the treatment group did not sign up. Indeed the operator had difficulties first to contact eligible welfare recipients, and then to convince them to enter the counseling program. As a result, only 17% of the treatment group in waves 1 and 2 were actually treated at some point (Figures 1.2 and 1.3). This participation rate is slightly lower in wave 3, certainly because the experiment went to an end sooner

Table 1.1: Descriptive statistics: Control and treatment groups

	Control group	Treatment group	Difference
Observations	4430	13236	
Proportion in wave 1	0.85	0.85	-0.00047
	(0.36)	(0.36)	(0.0062)
Proportion in wave 2	0.089	0.089	-0.00032
	(0.29)	(0.28)	(0.0049)
Proportion in wave 3	0.063	0.063	0.00078
	(0.24)	(0.24)	(0.0042)
Woman	0.46	0.46	-0.0028
	(0.5)	(0.5)	(0.0087)
$age \leq 30$	0.15	0.15	-0.0071
	(0.36)	(0.35)	(0.0062)
$30 < age \leq 40$	0.35	0.35	-0.0006
	(0.48)	(0.48)	(0.0083)
$40 < age \leq 50$	0.27	0.28	0.0093
	(0.44)	(0.45)	(0.0077)
$2 \leq seniority \leq 4$	0.53	0.5	-0.035*
	(1)	(0.99)	(0.018)
$4 < seniority \leq 6$	0.56	0.55	-0.012
	(0.5)	(0.5)	(0.0086)
Number of children	0.13	0.14	0.0089
	(0.34)	(0.35)	(0.0059)
Transfer 1 trimester	1117	1118	1.4
before entry	(428)	(439)	(7.5)
Transfer 2 trimesters	1090	1091	1.5
before entry	(442)	(458)	(7.7)
Transfer 3 trimesters	1074	1072	-1.6
before entry	(457)	(459)	(7.9)
Transfer 4 trimesters	740	735	-4.6
before entry	(342)	(345)	(5.9)
Employment rate 1 month	0.22	0.22	0.0019
before entry	(0.41)	(0.41)	(0.0071)
Employment rate 2 months	0.21	0.21	-0.00097
before entry	(0.41)	(0.41)	(0.007)

Scope: Individuals participating into the experiment. *: Significant coefficient at the 10% confidence level.

Figure 1.1: Cumulative distribution of seniority in wave 1



Note: Cumulative distribution of seniority into the welfare system, for the 14,980 individuals belonging to the first wave.

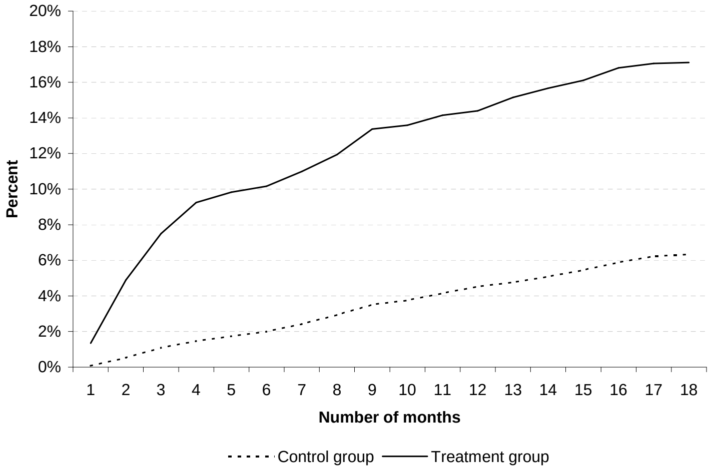
Lecture: Around 30% of individuals in wave 1 have less than 3 years of seniority into the welfare system.

for that wave (Figure 1.4).

The control group was supposed to be untreated: no advertising on the counseling scheme was targeted towards them, and the private operator was not supposed to contact them. However, some individuals did hear about the policy, and specifically asked to be included into the program. In that case, the private operator was allowed to enroll these individuals, even if they were part of the control group. This happened for about 6% of the control group in waves 1 and 2, and a similar pattern exists for wave 3.

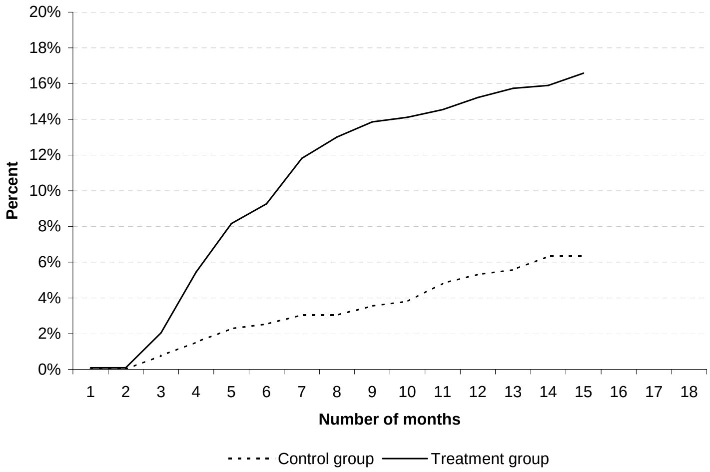
This low take-up, and the limited differential between individuals assigned to treatment and control group, is a source of low statistical precision in spite of the large sample size. Moreover, the fact that take-up is varying very strongly over time, as apparent from the figures, has important consequences for the identification strategy. The next section presents the outcomes and net cost parameters of interest and discusses identification in such a context.

Figure 1.2: Participation rate, wave 1



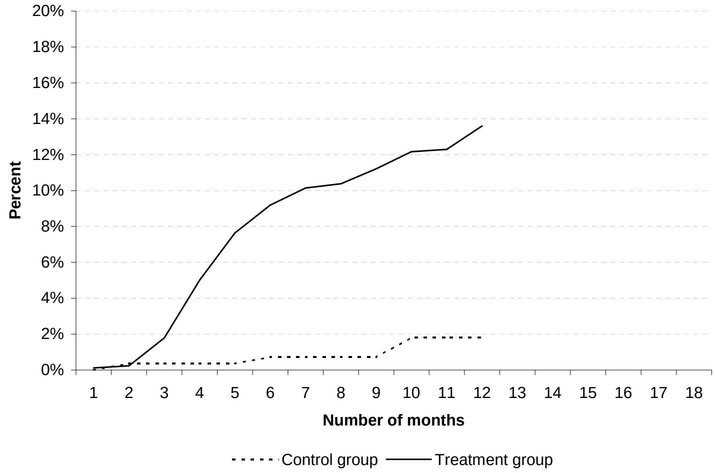
Note: Proportion of individuals in wave 1 that entered the couseling program, depending on whether they belonged to the treatment group (plain line) or the control group (dotted line).

Figure 1.3: Participation rate, wave 2



Note: Proportion of individuals in wave 2 that entered the couseling program, depending on whether they belonged to the treatment group (plain line) or the control group (dotted line).

Figure 1.4: Participation rate, wave 3



Note: Proportion of individuals in wave 3 that entered the couseling program, depending on whether they belonged to the treatment group (plain line) or the control group (dotted line).

1.3 A cost-benefit analysis

In the above design, there is one treatment that can have impact on three outcomes: the employment status of the beneficiary, e, the welfare transfer received by the beneficiary, paid on the public budget, T, and the price paid, also on the public budget, to the private provider of the policy, C. Both e and T can be defined on a monthly or quarterly basis, with calendar time indexed by t. For each period, we can define two counterfactuals for transfer and employment outcomes:

$$\left.\begin{array}{c} e_t^0 \\ T_t^0 \end{array}\right\} \text{ if the person has not entered treatment by time } t$$

$$\left.\begin{array}{c} e_t^1 \\ T_t^1 \end{array}\right\} \text{ if the person has entered treatment by time } t$$

It is clear that the counterfactual transfers received depend themselves on the chances of employment with and without treatment, but we don't need to make this structural relation explicit for the present purpose. The counterfactual price paid to the provider has two components: a lump-sum when the beneficiary enters treatment (A), and a premium when she gets a job during treatment window and keeps it a sufficient number of months (B). Call D the treatment window, that is the period during which the provider is in charge of the beneficiary and assume that a person entered treatment at date s. Whether the provider receives the premium B is a specific function of employment status over this period: call this function $I\left(\{e_t^1\}_{s\le t\le s+D}\right)$. Then, we can define the counterfactual payment to the provider:

$C_s^0 = 0$ if the person has not entered treatment by time s

$C_s^1 = A + B \times I\left(\{e_t^1\}_{s\le t\le s+D}\right)$ if the person has entered treatment at time s

If we restrict ourselves to the treatment window, we can compute the net public cost of sending someone into treatment, that is the price paid to the provider minus the possible cumulative gains in welfare payments:

$$C_s^1 - \sum_{s\le t\le s+D} \left(T_t^0 - T_t^1\right)$$

If the the policy is meant to be self-financed, then it should be continued only if this net public cost is zero or negative. This can be judged by testing:

$$E\left(C_s^1\right) \le \sum_{s\le t\le s+D} \left[E\left(T_t^0\right) - E\left(T_t^1\right)\right]$$

where the means can be conditioned on some relevant population. This is to say that the reduction in welfare transfers resulting from better access to jobs under the policy must fully compensate for the direct cost of it. Notice that it is not necessary to estimate separately the terms A and B in C^1 in order to compute the expected cost.

Otherwise, if the public decision maker is ready to pay in order to get welfare beneficiaries enter employment, then the counterpart to this cost is the change in employment status generated by the policy:

$$\sum_{s\le t\le s+D} \left(e_t^1 - e_t^0\right)$$

This is the policy causal impact on employment. Assume that the decision maker gives a monthly value W to the fact that a welfare recipient is employed rather than non-employed. Then, the policy is acceptable as long as:

$$W \cdot \left\{ \sum_{s\le t\le s+D} E\left(e_t^1\right) - E\left(e_t^0\right) \right\} \ge E\left(C_s^1\right) - \sum_{s\le t\le s+D} \left[E\left(T_t^0\right) - E\left(T_t^1\right)\right] \qquad (1.1)$$

Based on this, and provided that we can identify the counterfactual means contained in this expression, we can follow two routes: one is to compute the value W that can justify the policy, given the other parameters. We would however have a very imprecise estimation of W. Another is to fix W using alternative spending on employment policies or the market value of a job, or any other amount that can be considered a social value to having someone on the job, and test that the causal impact on employment, $\left(e_t^1 - e_t^0\right)$, is sufficiently large to justify the policy.

Notice that we will have little information on employment and welfare transfers effects outside the treatment window D, because our data has limited length. If one believes that treatment has lasting effects, then we may understate the returns to the policy. However, such lasting effects are rarely found in the literature (e.g. Card and Hyslop 2005). Moreover, our causal impacts will be small, so that this will not be a serious issue in practice.

1.4 Identification and estimation

Evaluating the general return to the policy along the above lines requires identification of the counterfactuals, e^0, e^1, T^0, T^1, C^1, at least for some population of interest. This is possible here in a very transparent way because the data is based on a random experiment. However, the date the treated enter treatment is not controlled: this dynamic structure of entry into treatment does restrict the model that can be identified. We can identify the model if there is either essential heterogeneity or if treatment effect is time-varying, but not both. This is a dichotomy that is present elsewhere in the evaluation literature based on duration models (Abbring and van den Berg 2003). We now discuss identification under several hypothesis.

Experiment starts when randomization has taken place and a given cohort of individuals may enter treatment any time. We use t for time elapsed since the start of the experiment. We note Z the random variable that takes value 1 if the individual lies on the list of persons that should be contacted in priority; $Z = 0$ otherwise. Z has been randomly determined and is constant over time. As we have previously shown, a significant number of persons did not comply with either treatment intention: few people from the list actually entered treatment during our period of observation and some people not on the list did enter treatment. Thus compliance appears to be dynamic in this framework: the same person, whatever its randomized status in the trial, may be untreated in t but treated in $t+1$, therefore a complier or a non-complier depending on time.

Define x_t an indicator variable that takes value 1 if an individual has entered treatment by time t since the start of the experiment (and zero otherwise). We thus have $E(x_t) = p(t)$, where $p(t)$ is the cumulative function of entry into treatment.

In this section, we use y as a generic outcome variable, representing either e, Tor C^1. Counterfactual outcomes with or without treatment respectively are $y^1_{t,s}$ and y^0_t, where t is time elapsed since randomization and s is the date at which individual actual entry into treatment took place. We will consider two cases. In this section and section 1.4.1, we first assume that counterfactual output $y^1_{t,s}$ (thus also impact) does *not* depend on s, ie length of exposure to treatment. In sections 1.4.2 and 1.4.3 we then assume that it does depend on s.

First assume, therefore, $y^1_{t,s} = y^1_t$. By definition of the counterfactuals, we observe:

$$y_t = x_t y^1_t + (1 - x_t) y^0_t$$

We can always assume that Z is independent from $\{y^1_t, y^0_t, x_t(Z{=}0), x_t(Z{=}1)\}$.

We have:

$$E(y_t|Z) = E(y^0_t|Z) + E\left(x_t(y^1_t - y^0_t)|Z\right)$$

If we assume that the distribution of treatment effect, $\{y^1_t - y^0_t\}$ is independent from the distribution of entry into treatment $\{x_t\}$[4], this simplifies to:

$$\begin{aligned} E(y_t|Z) &= E(y^0_t) + E(y^1_t - y^0_t)E(x_t|Z) \\ &= E(y^0_t) + E(y^1_t - y^0_t)p(t|Z) \end{aligned}$$

Therefore, we have the following result:

[4]Notice that the fact that y^1_t is not indexed on s does not imply that treatment effect could not be correlated with date of entry into treatment in the population, in the sense that those with high potential treatment impact tend to enter sooner (or the reverse).

Result 1

If counterfactual y_t^1 is not indexed on the date of entry into treatment s, and the date of entry into treatment is not correlated with the program impact $y_t^1 - y_t^0$, then the Wald estimator measures the average program impact at date t:

$$\frac{E(y_t|Z=1)-E(y_t|Z=0)}{p(t|Z=1)-p(t|Z=0)} \;=\; E(y_t^1 - y_t^0)$$

This estimator depends on t only through the time dependance of y_t^0 and y_t^1.

1.4.1 Essential heterogeneity

The modern evaluation literature is careful to the fact that entry into treatment may not be unrelated to the benefit from treatment. As a result, $\{y_t^1 - y_t^0\}$ may not be independent from $\{x_t\}$, a situation labeled "essential heterogeneity". In that case, we can identify a Local Average Treatment Effect in the sense of Angrist and Imbens (1994) if we assume a monotonicity assumption. Call $x_t(Z)$ the counterfactual treatment situation of any individual at time t depending on the intention to treat variable Z. We assume:

$$x_t(Z=1) \;\geq\; x_t(Z=0) \qquad \forall t$$

This means that an individual that was not on the priority list who happens to be treated however, would also be treated if he were on the priority list. In that case, using independence of counterfactuals with respect to Z, we have:

$$\begin{aligned}
&E(y_t|Z=1)-E(y_t|Z=0)\\
&= E(x_t(y_t^1-y_t^0)|Z=1)-E(x_t(y_t^1-y_t^0)|Z=0)\\
&= E(x_t(Z=1)(y_t^1-y_t^0))-E(x_t(Z=0)(y_t^1-y_t^0))\\
&= E([x_t(Z=1)-x_t(Z=0)]\,(y_t^1-y_t^0))\\
&= E((y_t^1-y_t^0)|\,[x_t(Z=1)-x_t(Z=0)]=1)\;P([x_t(Z=1)-x_t(Z=0)]=1)\\
&\quad +0\times P([x_t(Z=1)-x_t(Z=0)]=0)\\
&\quad +E(-(y_t^1-y_t^0)|\,[x_t(Z=1)-x_t(Z=0)]=-1)\;P([x_t(Z=1)-x_t(Z=0)]=-1)
\end{aligned}$$

Using monotonicity we have the following result:

Result 2
If counterfactual y_t^1 is not indexed on the date of entry into treatment s; if the date of entry into treatment can be correlated with the program impact $y_t^1 - y_t^0$; and under the above monotonicity assumption, the Wald estimator measures the average program impact at date t on the complier population at that date:

$$\frac{E(y_t|Z=1)-E(y_t|Z=0)}{p(t|Z=1)-p(t|Z=0)} = E\left[(y_t^1 - y_t^0)|x_t(Z=1)=1, x_t(Z=0)=0\right]$$

This estimator depends on t through the time dependance of y_t^0 and y_t^1 and changes in the complier population.

At any time t, treatment effect is identified over the population of compliers, as is necessarily the case under essential heterogeneity. Notice, however, that "complier" is a changing status. The same person can be a complier at t (if $x_t(Z=1)=1$, $x_t(Z=0)=0$), an always taker at $t+1$ (if $x_{t+1}(Z=1)=1$, $x_{t+1}(Z=0)=1$) and a never taker at $t-1$ (if $x_{t-1}(Z=1)=0$, $x_{t-1}(Z=0)=0$). The parameter is thus not identified over a stable population, and it can be time-varying for this reason.

1.4.2 Variable effect and no essential heterogeneity

In the above model, the policy impact, $y_t^1 - y_t^0$, is not indexed on the time elapsed *since entry into treatment.* We now consider this important possibility. We allow counterfactual $y_{t,s}^1$ to actually depend on s and we assume that if the person entered treatment at time s, then treatment effect at time $t \geq s$ is $y_{t,s}^1 - y_t^0 = \delta_{t-s}$. We thus assume that the effect of program participation now depends on the length of exposure to treatment $t-s$, instead of depending on t as in the previous section. In that case:

$$y_t = y_t^0 + \sum_{s=1}^{t} \delta_{t-s}\left[x_s - x_{s-1}\right]$$

(with $x_0 \equiv 0$).

If we assume that there is no essential heterogeneity, δ_{t-s} is independent from $\{x_t\}$ and we can write:

$$
\begin{aligned}
E(y_t|Z) &= E(y_t^0) + \sum_{s=1}^{t} E(\delta_{t-s}) E([x_s - x_{s-1}] \,|Z) \\
&= E(y_t^0) + \sum_{s=1}^{t} E(\delta_{t-s}) \left[p(s|Z) - p(s-1|Z)\right] \\
&= E(y_t^0) + \sum_{s=1}^{t} E(\delta_{t-s}) f(s|Z)
\end{aligned}
$$

where $f(s|Z) = [p(s|Z) - p(s-1|Z)]$ is the discrete-time equivalent to a density of entry into treatment. The whole sequence of $E(\delta_{t-s})$ is thus identified. For instance:

$$
E(y_1|Z=1) - E(y_1|Z=0) = E(\delta_0) \times [f(1|Z=1) - f(1|Z=0)]
$$

identifies $E(\delta_0)$. Then:

$$
\begin{aligned}
E(y_2|Z=1) - E(y_2|Z=0) = \;& E(\delta_0) \times [f(2|Z=1) - f(2|Z=0)] \\
& + E(\delta_1) \times [f(1|Z=1) - f(1|Z=0)]
\end{aligned}
$$

identifies $E(\delta_1)$, and so on. Therefore, we have:

Result 3

If counterfactual $y_{t,s}^1$ is indexed on the date of entry into treatment s such that impact is $y_{t,s}^1 - y_t^0 = \delta_{t-s}$; and if the date of entry into treatment is independent from the program impact, then it is possible to identify the set of parameters $E(\delta_\tau)$ for all observable values of τ. The estimator is an iterative version of the Wald estimator.

This estimator depends only on exposure to treatment $t-s$.

1.4.3 Variable effect and essential heterogeneity

If essential heterogeneity is present in addition to variable treatment effect, we can no longer identify a meaningful parameter, even under a monotonicity assumption.

Using the notation $\Delta x_t = [x_s - x_{s-1}]$, we have:

$$\begin{aligned}
&E(y_t|Z=1) - E(y_t|Z=0) \\
&\quad = \sum_{s=1}^{t} E(\delta_{t-s}\left[x_s - x_{s-1}\right]|Z=1) - E(\delta_{t-s}\left[x_s - x_{s-1}\right]|Z=0) \\
&\quad = \sum_{s=1}^{t} E(\Delta x_s(Z=1)\delta_{t-s}) - E(\Delta x_s(Z=0)\delta_{t-s}) \\
&\quad = \sum_{s=1}^{t} E(\delta_{t-s}\left[\Delta x_s(Z=1) - \Delta x_s(Z=0)\right])
\end{aligned}$$

Monotonicity is not sufficient because the two following situations are compatible with monotonicity:

1. $\Delta x_s(Z=1) = 1$ and $\Delta x_s(Z=0) = 0$, which is true for people who are never takers in $s-1$ and compliers in s;
2. $\Delta x_s(Z=1) = 0$ and $\Delta x_s(Z=0) = 1$, which is true for people who are compliers in $s-1$ and always takers in s.

Therefore:

Result 4

If essential heterogeneity is present in addition to variable treatment effect, we can no longer identify a meaningful parameter, even under a monotonicity assumption.

The general intuition for this negative result is that there are now two reasons why the estimated impact could vary with time: the population over which it is identified (the compliers) vary over time *and* the structural parameter varies with time. These two sources cannot be disentangled.

Abbring and van den Berg (2005) analyze some identification issues in social experiments with instrumental variable and duration outcomes. Their analysis of the dynamic take-up issue refers to a different setup than this one. They emphasize in particular that knowledge of their assignment status is likely to influence employment behavior of individuals even before they actually enter treatment. This generates a direct effect of the instrument on the outcome, that limits substantially identification. This does not seem relevant in our context, because people are unaware of their assignment status and not even aware that there is something as an assignment status. The restrictions to identification that we discuss here come from a different source, not examined by Abbring and van den Berg (2005), namely essential heterogeneity.

In practice, we will provide the two estimations, and they will prove to be hardly different, so that the distinction is one of principle in this application. In theory, though, it might be better to think in terms of a model with time-varying effects. There are two reasons for this. First, counseling is viewed by providers as a process that requires time to gain efficiency. In contrast, we cannot judge the extent to which self-selection into treatment could be based on actual treatment impact (especially given that the treatment is new and individuals have little way to make an appropriate judgment on its efficiency). Second, evaluation of equation (1.1) requires that treatment impact is measured over the same population at consecutive dates. This is not possible under essential heterogeneity.

1.4.4 Estimations

Estimations are run on a quarterly basis, and we have information on 4, 5 or 6 successive trimesters after entry into the experiment depending on the individual (see section 1.2.2). We set employment status e_t at trimester t equal to the number of months one has been employed during that trimester. So e_t takes values in $\{0, 1, 2, 3\}$. T_t is the amount of welfare transfer received during trimester t by the individual, and x_t is a dummy set to 1 if the individual has entered treatment by date t.

All the estimators presented in the previous section use combinations of reduced form (or "intention to treat") parameters, as does the Wald estimator. We thus estimate a set of reduced form parameters that are then combined according to the above formulas, in order to form the structural parameters of interest that will be shown in the tables.

Estimations are thus run in two successive steps. We first simultaneously estimate the following OLS regressions :

$$y = \alpha \ + \beta \ Z + \epsilon \tag{1.2}$$

where y can either be employment status e_t, welfare transfer T_t, cost C_t^1 or treatment status x_t at date t ($1 \leq t \leq 6$). $\hat{\beta}$ gives an estimation of $[E(y|Z{=}1){-}E(y|Z{=}0)]$ for each of these 24 dependant variables (4 different outcomes and 6 time periods), along with a variance-covariance matrix robust to heteroscedasticity.

Sections 1.4.1 and 1.4.2 showed that the parameters of interest under essential heterogeneity and variable effect are functions of these 24 coefficients. So we use them to compute estimates of the causal impacts and their standard errors (using the delta method) for the 6 trimesters under both frameworks.

In order to improve efficiency, we also estimate the system of equations (1.2) with a set of additional covariates : age, sex, number of children, city of residence, amount of welfare transfer received in each of the 4 trimesters before entry into the experiment, and seniority into the welfare benefit system.

Eventually, we implement the test described in (1.1) to check whether the policy is profitable for the public budget. We compute

$$R = W \cdot \left\{ \sum_{s \leq t \leq s+D} E\left(e_t^1\right) - E\left(e_t^0\right) \right\} - E\left(C_s^1\right) + \sum_{s \leq t \leq s+D} \left[E\left(T_t^0\right) - E\left(T_t^1\right)\right]$$

for given values of W. The policy is considered to be profitable if R is statistically greater than 0. The variance of R is computed using the variance-covariance matrix of the simultaneously estimated causal effects on employment, transfers and costs at all relevant trimesters, obtained using equation (1.2).

In our context, three possibilities naturally arise when determining the value of W that measures the value that society or the administration gives to having someone on the job rather than not. The first one is the wage for the kind of job that this population is likely to enter, equal to the minimum wage (around 1,000 € per month in employer cost). An alternative is the public cost of subsidized jobs targeted towards welfare recipients (*CIRMA* or *CAV*). This is a measure of the amount of money that is spent on alternative active employment policies to have low educated people on a job. Depending on the type of contract, the amount spent on public funds varies from around 450 € per month to 90% of the minimum wage. Eventually, we can focus exclusively on the district budget instead of society welfare, and therefore set W to 0 €. In total, we compute R for $W = 0$ € , $W = 450$ € and $W = 1,000$ €.

1.5 Results

Table 1.2 shows the results of the estimations on the whole sample. The coefficient at the intersection of row 'Trimester n' and column 'essential heterogeneity' is the LATE estimated on compliers of the n^{th} trimester under essential heterogeneity. The interpretation is different under a variable effect framework (the last two columns): the coefficient is the causal impact during the n^{th} trimester after entry into treatment.

The employment variable is based on a reporting card that each welfare beneficiary must send every trimester at some anniversary date. No adjustment is made for part-time employment (on which we have no reliable information): as our employment social value W is based on full-time work, we could overstate the value we give to employment if treatment induces more full-time jobs; but also understate it if it induces more part-time jobs. This does not affect the purely financial net cost evaluation, however (when $W = 0$). When individuals get out of welfare, we no longer have a reporting card: we assume that they are employed and we know they no longer receive transfers.

The amount of transfers is directly dependant on work income. Individuals with no income receive the full amount of transfer, that varies with family composition. As they earn more income, their amount of transfer is reduced (in most cases at a 50% implicit tax rate, but the details are more complex) up to a point where they are out of the program (for a single person with no child, this is about when she earns the full time minimum wage). It is important to note that transfers paid a given trimester are based on income reported a trimester earlier. This generates a lag between our employment variable and our transfer variable.

A similar pattern emerges in Table 1.2 for both essential heterogeneity and variable effects frameworks, whether or not covariates are included: the counseling program often has a positive impact on employment, but this effect is not significant at any date. A coefficient of 0.17 on trimester 2, for instance, means that an additional 0.17 month is worked over the trimester as a result of treatment. This seems a low figure, but the control group worked only 0.65 month during the first trimester[5], so the impact is relatively large in proportion. Unfortunately, because of low take-up, this is very imprecisely estimated.

Treatment effect on welfare transfers is always negative, but significantly different from 0 only on the third trimester. It is very unlikely that the counseling scheme should affect transfers otherwise than through employment (the only other determinant of transfer is household composition). Therefore, impacts on those two outcomes should be linked. Most of the employment impact seems to happen during the second and third trimesters, as if it took the first trimester for the counselor to prepare some action and set up a search program for the person. This same timing is apparent on the transfer variable. Remember that there is a trimester lag between employment and transfer, and that treatment can start at any date between two successive income reporting: as such, trimester 1 coefficients for transfer are only partly affected by the program. Indeed, the point estimates, although negative, are much smaller for this first trimester than later on. Thereafter, the shape of transfer impacts are related with the shape of employment impacts (with a lag). At trimester 3, the significant transfer impact is rather large at around 200 € , representing about 20% of an average baseline transfer.

The average amount of money received per treated by the provider is close to 3,150 € in all specifications (see section 1.2.1 for a description of the payment scheme). The cost benefit analysis reveals a clear deficit for the public budget. This was expected, given that the average welfare payment is about 1,000 € per trimester and the program only has limited impact on employment and resulting transfers.

[5] As apparent from Figures 1.2-1.4, only a very small share of the control group entered treatment during the first trimester of the experiment, so that this value is close to the counterfactual employment rate under no treatment for the whole population.

Interestingly enough, the net public cost can be estimated quite precisely in spite of the large standard errors on most of the transfer and employment impacts. This is because the cost component is high enough to cover many likely values of the other parameters.

However this last result must be interpreted with caution in Table 1.2, because the costs and benefits are not estimated on a stable population for all trimesters. As explained in section 1.2.2, we lack information on wave 2 for trimester 6 and on wave 3 for trimesters 5 and 6. Therefore estimations in Table 1.2 are based on a stable population only for the first four trimesters. A solution to this problem is to run the estimations only on individuals present during the 6 trimesters. Table 1.3 shows the results of such estimations.

Based on 85% of the total sample (wave 1 individuals), results in Table 1.3 are similar to those on the whole population. Although the impact on employment is positive and greater than previously during trimesters 1 and 2, it is not significant at the 10% confidence level. The negative and significant impact on welfare transfers in the third trimester is still present, and is even stronger than in Table 1.2; effects at trimesters 2 and 4 are also stronger and more significant, in line with the higher employment point estimates at trimesters 1 and 3. The cost benefit analysis is now meaningful, since all individuals are present over the whole experiment. Over this observation window, payment to the counseling firm has been considerably higher than gains in welfare transfers, by more than 2,200 € . Thus the counseling scheme is far from cost-effective for the public budget. Because employment intensity is limited, this is not counterbalanced by a monthly social value of $W = 450$ € for putting a welfare recipient to work. Remember this is the subsidy paid by the State on a class of jobs targeted towards welfare recipients. With this social value to employment, net cost remains strongly negative and significantly so. When the value of work is set to about the minimum wage employer cost $W = 1,000$ €, the effect is still negative but not significant anymore, certainly because this specification gives more weight to the - imprecisely estimated - impact on employment. Also, this value for W assumes full time work versus no work, which is probably not the general case because of the incidence of part-time work (remember we do

Table 1.2: Causal impact of counseling, whole population

	Model hypothesis:			
	Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	0.037	0.04	0.037	0.04
	(0.38)	(0.37)	(0.38)	(0.37)
Trimester 2	0.13	0.13	0.17	0.17
	(0.27)	(0.26)	(0.25)	(0.24)
Trimester 3	0.084	0.08	0.06	0.055
	(0.22)	(0.22)	(0.23)	(0.22)
Trimester 4	-0.0055	-0.0067	-0.094	-0.094
	(0.22)	(0.22)	(0.26)	(0.26)
Trimester 5	0.02	0.014	0.051	0.042
	(0.21)	(0.2)	(0.24)	(0.24)
Trimester 6	0.0087	0.017	0.0054	0.026
	(0.19)	(0.19)	(0.22)	(0.22)
Welfare transfer				
Trimester 1	-75	-73	-75	-73
	(139)	(91)	(139)	(91)
Trimester 2	-136	-131	-162	-156
	(110)	(87)	(113)	(100)
Trimester 3	-195**	-192**	-249**	-249**
	(99)	(83)	(111)	(105)
Trimester 4	-104	-106	-23	-27
	(105)	(92)	(127)	(114)
Trimester 5	-105	-95	-95	-75
	(99)	(88)	(118)	(111)
Trimester 6	-108	-107	-135	-141
	(96)	(86)	(113)	(107)
Cost				
	3145***	3152***	3147***	3155***
	(43)	(44)	(48)	(49)
Cost - Benefit				
W= 0 €	-2422***	-2448***	-2407***	-2435***
	(556)	(425)	(535)	(432)
W= 450 €	-2297**	-2324***	-2301**	-2327***
	(949)	(806)	(892)	(773)
W=1,000 €	-2144	-2172	-2172	-2194
	(1612)	(1460)	(1495)	(1364)
Covariates	no	yes	no	yes

Scope: 17,666 individuals participating into the experiment. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

Table 1.3: Causal impact of counseling, wave 1

	Model hypothesis:			
	Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	0.12	0.12	0.12	0.12
	(0.35)	(0.34)	(0.35)	(0.34)
Trimester 2	0.26	0.26	0.3	0.3
	(0.28)	(0.28)	(0.28)	(0.27)
Trimester 3	0.13	0.12	0.079	0.067
	(0.24)	(0.23)	(0.24)	(0.24)
Trimester 4	0.017	0.015	-0.075	-0.076
	(0.24)	(0.24)	(0.27)	(0.26)
Trimester 5	-0.083	-0.083	-0.15	-0.15
	(0.23)	(0.22)	(0.25)	(0.25)
Trimester 6	0.0087	0.017	0.037	0.051
	(0.23)	(0.22)	(0.25)	(0.24)
Welfare transfer				
Trimester 1	-47	-27	-47	-27
	(132)	(77)	(132)	(77)
Trimester 2	-181	-169*	-218*	-208**
	(118)	(89)	(122)	(99)
Trimester 3	-282***	-274***	-362***	-360***
	(108)	(89)	(117)	(107)
Trimester 4	-203*	-200**	-156	-155
	(115)	(100)	(131)	(117)
Trimester 5	-128	-124	-67	-64
	(111)	(98)	(126)	(118)
Trimester 6	-108	-107	-94	-94
	(113)	(101)	(127)	(119)
Cost				
	3145***	3152***	3146***	3153***
	(51)	(51)	(55)	(55)
Cost - Benefit				
W= 0 €	-2197***	-2251***	-2203***	-2246***
	(608)	(457)	(604)	(481)
W= 450 €	-1996*	-2050**	-2065**	-2105**
	(1026)	(865)	(1002)	(863)
W=1,000 €	-1750	-1803	-1896	-1933
	(1734)	(1567)	(1676)	(1527)
Covariates	no	yes	no	yes

Scope: 14,980 individuals participating into the first wave of the experiment. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

not observe working time). Generally, therefore, although the employment impacts of the program are imprecisely estimated, due to the low take up rate into the program, its cost-effectiveness is very clearly established by the data and this results from the high price paid to the counseling firm. The fact that more than half of the average cost is lump-sum payment, received for every person entering the program, certainly plays a strong role.

Tables 1.4 to 1.7 are variants of the previous estimations on wave 1, where we focus on specific subpopulations based on age, tenure into the welfare system, and children. The general picture is not affected by age or the presence of children in the household (Tables 1.4, 1.5 and 1.7), although the cost-benefit analysis is less negative in the presence of children. This is because welfare payments are much higher, so that decreases in transfer, when they occur, are also substantially stronger.

Table 1.6 illustrates the presence of a strong impact heterogeneity. Obviously, a counseling policy is much more efficient on a population that does not have too much tenure in welfare. When we consider only individuals with 2 to 4 years of seniority (as opposed to more than 4), we find stronger and longer lasting employment and transfer impacts. In particular, transfer impacts are stable and statistically significant all over the period. It could imply that this population is closer to the labor market and more likely to find a job and remain employed with some outside help. The cost/benefit analysis still reveals a deficit for the public budget, but it is now estimated at 1,700 €, thanks to the stronger reductions in welfare transfer.[6] When employment is given a social value, a zero net cost becomes more likely. This implies that such policies may be efficient and cost effective for some populations and not for others, something that can only be judged by systematic evaluation. It questions the policy decision to target this type of policy to individuals with long tenure in welfare, even if this population is of strong need of help.

[6] This refers to estimates with control variable, that are not significantly different from the other estimates, but are more precise.

Table 1.4: Causal impact of counseling, wave 1, less than 40 years old

	Model hypothesis:			
	Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	0.42	0.57	0.42	0.57
	(0.57)	(0.56)	(0.57)	(0.56)
Trimester 2	0.47	0.56	0.49	0.56
	(0.43)	(0.42)	(0.41)	(0.4)
Trimester 3	0.26	0.29	0.12	0.087
	(0.35)	(0.34)	(0.36)	(0.35)
Trimester 4	0.17	0.2	0.072	0.11
	(0.35)	(0.34)	(0.41)	(0.4)
Trimester 5	0.089	0.13	0.045	0.095
	(0.34)	(0.33)	(0.38)	(0.38)
Trimester 6	0.092	0.15	0.073	0.14
	(0.34)	(0.33)	(0.38)	(0.37)
Welfare transfer				
Trimester 1	145	-41	145	-41
	(195)	(126)	(195)	(126)
Trimester 2	-73	-177	-153	-228
	(169)	(137)	(178)	(158)
Trimester 3	-207	-275**	-350**	-377**
	(149)	(129)	(167)	(161)
Trimester 4	-176	-240*	-125	-195
	(158)	(142)	(188)	(173)
Trimester 5	-121	-179	-67	-116
	(155)	(141)	(185)	(176)
Trimester 6	-90	-147	-75	-130
	(159)	(147)	(184)	(175)
Cost				
	3278***	3276***	3282***	3280***
	(76)	(77)	(84)	(85)
Cost - Benefit				
W= 0 €	-2756***	-2217***	-2656***	-2194***
	(849)	(680)	(840)	(707)
W= 450 €	-2084	-1365	-2110	-1493
	(1510)	(1311)	(1460)	(1293)
W=1,000 €	-1262	-324	-1444	-636
	(2594)	(2370)	(2477)	(2283)
Covariates	no	yes	no	yes

Scope: 7,124 individuals participating into the first wave of the experiment, less than 40 years old. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

Table 1.5: Causal impact of counseling, wave 1, more than 40 years old

	Model hypothesis:			
	Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	-0.06	-0.23	-0.06	-0.23
	(0.44)	(0.43)	(0.44)	(0.43)
Trimester 2	0.14	-0.0087	0.17	0.035
	(0.37)	(0.36)	(0.37)	(0.36)
Trimester 3	0.059	-0.074	0.06	-0.063
	(0.32)	(0.32)	(0.33)	(0.32)
Trimester 4	-0.067	-0.2	-0.14	-0.27
	(0.33)	(0.32)	(0.35)	(0.35)
Trimester 5	-0.19	-0.29	-0.26	-0.35
	(0.31)	(0.31)	(0.34)	(0.34)
Trimester 6	-0.0047	-0.094	0.05	-0.019
	(0.3)	(0.3)	(0.32)	(0.32)
Welfare transfer				
Trimester 1	-226	19	-226	19
	(179)	(94)	(179)	(94)
Trimester 2	-314*	-130	-331**	-159
	(165)	(112)	(167)	(121)
Trimester 3	-391**	-246**	-435***	-318**
	(155)	(120)	(164)	(141)
Trimester 4	-269	-137	-224	-94
	(167)	(139)	(183)	(159)
Trimester 5	-173	-59	-105	-9.7
	(156)	(136)	(174)	(160)
Trimester 6	-166	-50	-144	-42
	(157)	(140)	(175)	(163)
Cost				
	3024***	3032***	3028***	3038***
	(67)	(68)	(71)	(72)
Cost - Benefit				
W= 0 €	-1485*	-2430***	-1563*	-2435***
	(862)	(610)	(859)	(646)
W= 450 €	-1541	-2835**	-1642	-2839**
	(1375)	(1137)	(1351)	(1145)
W=1,000 €	-1610	-3331	-1740	-3334
	(2291)	(2068)	(2232)	(2037)
Covariates	no	yes	no	yes

Scope: 7,856 individuals participating into the first wave of the experiment, more than 40 years old. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

Table 1.6: Causal impact of counseling, wave 1, between 2 and 4 years of seniority

	Model hypothesis: Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	0.6	0.54	0.6	0.54
	(0.51)	(0.5)	(0.51)	(0.5)
Trimester 2	0.55	0.48	0.53	0.47
	(0.38)	(0.38)	(0.37)	(0.37)
Trimester 3	0.32	0.24	0.16	0.066
	(0.32)	(0.32)	(0.33)	(0.33)
Trimester 4	0.14	0.074	0.012	-0.047
	(0.32)	(0.32)	(0.37)	(0.37)
Trimester 5	0.093	0.032	0.039	-0.015
	(0.31)	(0.3)	(0.34)	(0.34)
Trimester 6	0.12	0.066	0.12	0.074
	(0.3)	(0.3)	(0.34)	(0.34)
Welfare transfer				
Trimester 1	-246	-123	-246	-123
	(179)	(110)	(179)	(110)
Trimester 2	-333**	-239*	-362**	-277**
	(156)	(124)	(163)	(141)
Trimester 3	-354**	-278**	-385**	-326**
	(140)	(121)	(155)	(149)
Trimester 4	-329**	-258*	-301*	-229
	(150)	(134)	(175)	(160)
Trimester 5	-349**	-279**	-367**	-300*
	(145)	(131)	(165)	(156)
Trimester 6	-358**	-298**	-374**	-324**
	(147)	(136)	(166)	(158)
Cost				
	3280***	3288***	3280***	3287***
	(74)	(75)	(80)	(82)
Cost - Benefit				
W= 0 €	-1311*	-1813***	-1246	-1708***
	(791)	(623)	(783)	(650)
W= 450 €	-491	-1171	-587	-1222
	(1376)	(1206)	(1332)	(1192)
W=1,000 €	511	-385	219	-627
	(2339)	(2175)	(2241)	(2104)
Covariates	no	yes	no	yes

Scope: 7,302 individuals participating into the first wave of the experiment, between 2 and 4 years of seniority into the welfare system. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

Table 1.7: Causal impact of counseling, wave 1, with at least one child

	Model hypothesis:			
	Essential heterogeneity		Variable effect	
	(1)	(2)	(3)	(4)
Employment				
Trimester 1	0.097	0.0092	0.097	0.0092
	(0.6)	(0.57)	(0.6)	(0.57)
Trimester 2	0.39	0.29	0.46	0.36
	(0.49)	(0.47)	(0.49)	(0.47)
Trimester 3	0.36	0.27	0.42	0.33
	(0.4)	(0.4)	(0.42)	(0.41)
Trimester 4	0.086	0.019	-0.11	-0.15
	(0.41)	(0.4)	(0.46)	(0.45)
Trimester 5	0.03	-0.04	-0.071	-0.13
	(0.41)	(0.4)	(0.47)	(0.46)
Trimester 6	-0.036	-0.1	-0.037	-0.1
	(0.41)	(0.41)	(0.46)	(0.46)
Welfare transfer				
Trimester 1	-47	21	-47	21
	(280)	(122)	(280)	(122)
Trimester 2	-280	-221	-335	-276
	(247)	(156)	(250)	(176)
Trimester 3	-490**	-439***	-665***	-611***
	(219)	(168)	(239)	(216)
Trimester 4	-447*	-407**	-439	-405*
	(237)	(197)	(273)	(241)
Trimester 5	-367	-335	-274	-256
	(237)	(205)	(275)	(246)
Trimester 6	-333	-318	-305	-301
	(247)	(220)	(275)	(250)
Cost				
	3289***	3311***	3304***	3326***
	(110)	(114)	(115)	(118)
Cost - Benefit				
W= 0 €	-1325	-1613*	-1239	-1498
	(1271)	(849)	(1269)	(916)
W= 450 €	-910	-1409	-899	-1359
	(1938)	(1520)	(1916)	(1558)
W=1,000 €	-401	-1161	-484	-1189
	(3097)	(2696)	(3037)	(2690)
Covariates	no	yes	no	yes

Scope: 4,105 individuals participating into the first wave of the experiment, with at least one child. *, **, ***: Significant coefficient at the 10%, 5% and 1% confidence level. See sections 1.4.1 and 1.4.2 for estimation formula under each model hypothesis. W is a monthly social value for employment.

1.6 Conclusion

This paper studies a job-search counseling program targeted at individuals that have been on welfare for at least two years. This population has severe difficulties to access the labor market, and was offered the possibility to benefit from intensive counseling. This program was provided by a private operator, and was costly to the public budget.

We evaluate the impact of this policy on job access and its cost-effectiveness, using a random allocation of welfare recipients to the program. Although the low take-up of the program implies that some parameters are imprecisely estimated, we can reach strong conclusions of interest for public policy. First, there seems to be some effects of the program on employment and resulting transfers. Second, those effects are, in general, far too small to ensure that the program is cost-effective, even if we give employment situations a social value on top of the transfer gains for the public budget. Third, there is significant heterogeneity, and we can isolate a population of relatively recent welfare status, for which the program is more efficient, still not cost-effective from a strictly financial point of view, but for which the net budget cost is closer to the order of magnitude of the social value of employment.

Therefore, the policy as implemented receives a poor evaluation. But one may wonder how general is this finding. Obviously, there is scope for more targeted intervention, and more experimentation would be needed to decide on which kind of population precisely. In particular, we cannot decide here if individuals with less than 2 years tenure in welfare, or more or less educated people, should benefit from job search counseling. Also, the pricing level and mechanism of the private counseling firm is the source of high costs. We don't know if more competition on that market could induce lower costs to the public budget or if the policy is intrinsically costly. Eventually, it is possible that the important lump-sum component of the cost delivers a poor incentive to the firm that may be a reason for its limited efficiency, especially in the long run. On that side too, more experimentation is needed.

Chapitre 2

L'impact sur les salaires de la durée d'une interruption de carrière suite à une naissance

Sommaire

Ce chapitre est accepté pour publication dans *Annales d'Economie et de Statistique*.

Après avoir donné naissance à un enfant, beaucoup de femmes se retirent du marché du travail plus longtemps que la durée légale du congé maternité. Cet article met en évidence que la durée de ce retrait temporaire a un impact négatif sur le salaire perçu après le retour à l'emploi, et quantifie l'ordre de grandeur de cet effet.

Il existe dans beaucoup de pays des dispositifs permettant aux jeunes parents de diminuer leur offre de travail dans les mois qui suivent la naissance d'un enfant. Dans la plupart des cas, un lien positif a été mis en évidence entre la période maximale d'indemnisation de ces dispositifs et la durée d'interruption de carrière observée après une naissance chez les jeunes parents éligibles (Lalive and Zweimüller 2005, Piketty 2005, Del Boca and Wetzels 2008). Il semble donc que de telles politiques familiales aient un impact fort sur le taux d'emploi des parents. En revanche, la littérature est moins développée concernant l'impact que ces politiques pourraient avoir sur le déroulement de la carrière, une fois le retour à l'emploi effectué. Concernant les arrêts de carrière consécutifs à une naissance, les articles se focalisent principalement sur l'écart de salaire pouvant exister suite à une naissance entre hommes et femmes (Gupta and Smith 2002, Hotchkiss and Pitts 2007), ou entre femmes avec et sans enfant (Buding and England 2001, Davies and Pierre 2005, Felfe 2006). Mais le plus souvent, ils ne différencient pas l'effet sur les salaires de la naissance proprement dite de celui de la durée d'interruption. Ruhm (1998) fait cette distinction et utilise la création de ces politiques familiales dans plusieurs pays européens entre 1969 et 1993 pour montrer que les salaires diminuent avec la longueur de la période d'interruption. Plus récemment plusieurs articles mettent en évidence sur données allemandes qu'une année d'interruption diminuerait le salaire de 6% à 20% selon les spécifications (Ondrich et al. 2002, Kunze and Ejrnaes 2004, Beblo et al. 2006). La nature causale du lien entre durée d'interruption et salaire n'a pas encore été étudiée sur données françaises.

La durée de l'interruption de carrière n'est qu'une cause parmi d'autres pouvant expliquer la différence de salaire observée entre des femmes ayant des durées d'interruption différentes. Or certains des mécanismes en jeu ont également une influence sur le choix de la durée d'interruption (être plus ou moins carriériste, avoir de bonnes perspectives de carrière, etc.). Il n'est alors pas possible d'observer directement la part de la différence de salaire due à la durée de l'interruption. Une manière de résoudre ce problème d'endogénéité est de disposer d'une variable instrumentale. Nous utilisons ainsi une réforme de l'Allocation Parentale d'Éducation (APE) pour identifier l'effet causal de la durée d'interruption sur les salaires. Initialement ciblé sur les naissances de rang 3 et plus, ce dispositif a été étendu aux parents de deux enfants en juillet 1994. Piketty (2005) et Pailhé and Solaz (2006) observent que cette réforme a incité un grand nombre de femmes à diminuer leur offre de travail dans les trois années suivant la naissance de leur deuxième enfant. Le présent article prolonge ces deux études en s'intéressant aux évolutions de carrière une fois que le retour à l'emploi a été effectué. Les estimations en doubles ou triples différences sont basées sur l'appariement de l'Échantillon Démographique Permanent et des Déclarations Annuelles de Données Sociales. Les résultats obtenus suggèrent que la durée d'un retrait du marché du travail affecte négativement le salaire journalier dans les années qui suivent le retour à l'emploi, même si cette diminution n'est pas statistiquement significative dans toutes les spécifications retenues. Une partie de cette baisse pourrait être due à une diminution du nombre d'heures rémunérées par jour, hypothèse que les données ne nous permettent cependant pas de tester. Cette baisse de salaire ne semble pas être transitoire, car elle ne s'est pas encore estompée 10 ans après la naissance. Il est donc possible que la durée d'interruption continue à avoir un impact sur les revenus après la fin de la vie active, puisque le montant de la pension de retraite dépend directement des salaires perçus durant les 25 meilleures années de carrière dans le secteur privé.

2.1 Concilier vies professionnelle et familiale

2.1.1 Le système français

Plusieurs dispositifs ont été créés en France pour faciliter les interruptions d'activité des jeunes parents dans les années suivant la naissance d'un enfant. Ils ont été définis en vue de lever les deux principaux freins qui pourraient empêcher les parents de quitter leur travail pour élever leur enfant. D'une part les parents anticipent qu'il leur sera peut-être difficile de retrouver un emploi quand ils voudront revenir sur le marché du travail, et donc préfèrent ne pas quitter leur emploi actuel. D'autre part, diminuer son offre de travail signifie que les revenus du foyer diminuent, ce qui pèse sur la contrainte de budget des ménages. Deux politiques publiques, à mi-chemin entre politiques de l'emploi et politiques familiales, ont été mises en place pour contrer chacun de ces deux freins potentiels.

Le Congé Parental d'Éducation (CPE) permet aux parents qui le souhaitent d'arrêter de travailler suite à la naissance d'un enfant, en ayant l'assurance de retrouver chez leur ancien employeur un emploi équivalent à celui qu'ils quittent. Créé en 1977 pour les mères ayant au moins une année d'ancienneté chez cet employeur au moment de la naissance, il a été étendu aux pères avec la même condition d'éligibilité en 1984. Son but est de limiter les barrières au retour à l'emploi.

L'Allocation Parentale d'Éducation (APE) est une somme forfaitaire versée aux jeunes parents qui ne travaillent pas à temps plein, et ce jusqu'au troisième anniversaire de l'enfant. Cela concerne donc les travailleurs qui se retirent totalement du marché du travail, ceux qui passent de temps plein à temps partiel, et ceux qui étaient déjà à temps partiel avant la naissance et qui le restent. Enfin cela concerne aussi des personnes qui ne travaillent pas au moment de la naissance, mais qui satisfont néanmoins les critères d'éligibilité (par exemple, avoir travaillé 2 années dans les 5 années précédant la naissance d'un enfant de rang 2 postérieure à juillet 1994). Le montant de l'allocation varie suivant que le retrait du marché du travail est total ou partiel (493 euros pour l'APE à taux plein en 2003). L'APE permet donc de compenser une partie de la baisse de revenu due à la diminution de l'offre de travail. Créée en 1985 pour les parents d'un enfant de rang 3 ou plus, l'APE a été étendue en juillet 1994 aux naissances de rang 2[1].

Les personnes éligibles à ces deux dispositifs peuvent en bénéficier simultanément. Bien que ces politiques soient en théorie accessibles aussi bien aux pères

[1]Depuis 2004, elle est remplacée par le Complément de Libre Choix d'Activité, l'une des composantes de la Prestation d'Accueil du Jeune Enfant (Legendre et al. 2003). Les caractéristiques du CLCA sont très proches de celles de l'APE ; les deux principales différences sont une légère modification des conditions d'éligibilité et l'extension du dispositif aux naissances de rang 1 pour une durée maximale de 6 mois après la naissance. La création du CLCA en 2004 est mentionnée ici pour mémoire, car la présente étude porte sur des naissances ayant eu lieu avant 2004.

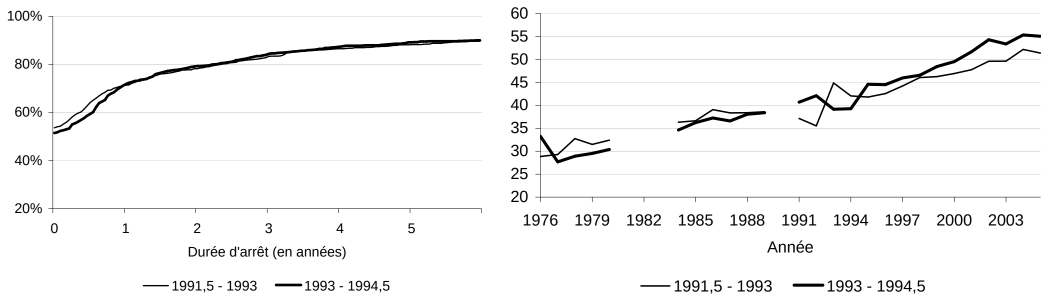
qu'aux mères, les femmes constituent la quasi-totalité des bénéficiaires (plus de 98% dans le cas de l'APE d'après Piketty 2005). C'est pourquoi cet article se concentre sur les évolutions salariales des femmes uniquement. En particulier, il n'est pas ici question d'étudier d'éventuelles inégalités salariales entre hommes et femmes. Il ne s'agit pas non plus de regarder si le simple fait d'avoir un enfant est pénalisant pour la suite de la carrière. Le point de vue adopté est de se restreindre aux femmes qui ont des enfants, et d'étudier si la durée de l'interruption de carrière consécutive à la naissance affecte la suite de leur carrière.

2.1.2 La réforme de l'APE de 1994

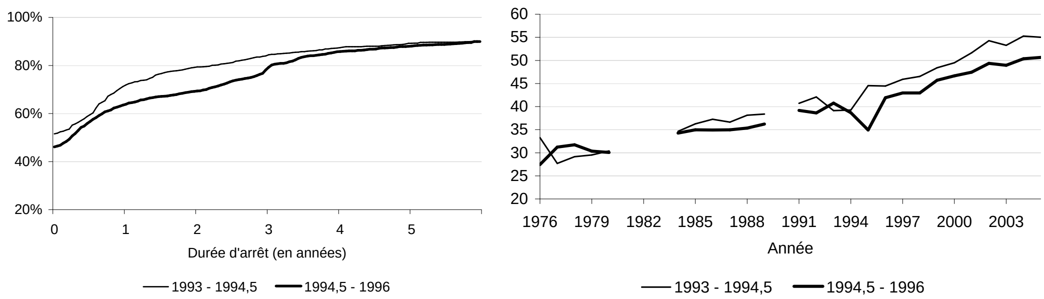
La réforme utilisée pour identifier l'effet de la durée sur les salaires est celle de l'APE en 1994. Jusqu'en juin 1994, l'APE n'était accessible que pour une naissance de rang 3 ou plus. Les parents d'un deuxième enfant ne bénéficiaient donc pas d'incitation particulière pour réduire temporairement leur offre de travail après la naissance. A partir de juillet 1994, l'APE leur permet de compenser partiellement la perte de revenu occasionnée par un retrait du marché du travail. Cela devrait donc se traduire par une augmentation du nombre de personnes ayant un enfant de rang 2 qui décident de quitter temporairement le marché du travail, ainsi que par un allongement de la durée de ce retrait (dans la limite de trois ans).

L'extension de l'APE a été un succès immédiat en terme de nombre de bénéficiaires (240 000 dès 1996) qui ne s'est pas démenti par la suite (plus de 315 000 en 2002). Cela ne garantit cependant pas que les interruptions de carrière ont été plus nombreuses et/ou plus longues, comme la théorie le prédit. Il est en effet possible que les bénéficiaires auraient de toutes façons réduit leur offre de travail, même en l'absence de compensation financière. Dans ce cas, l'incitation à modifier le comportement de participation au marché du travail n'aurait pas eu l'effet attendu, et le nombre élevé de bénéficiaires s'expliquerait principalement par un effet d'aubaine. Piketty (2005) et Pailhé and Solaz (2006) ont montré que cela n'est pas le cas, en se basant respectivement sur l'enquête Emploi et l'enquête Familles et Employeurs. Selon eux, l'extension de l'APE a eu un impact négatif important sur le taux d'emploi des femmes dans les trois ans suivant une naissance de rang 2.

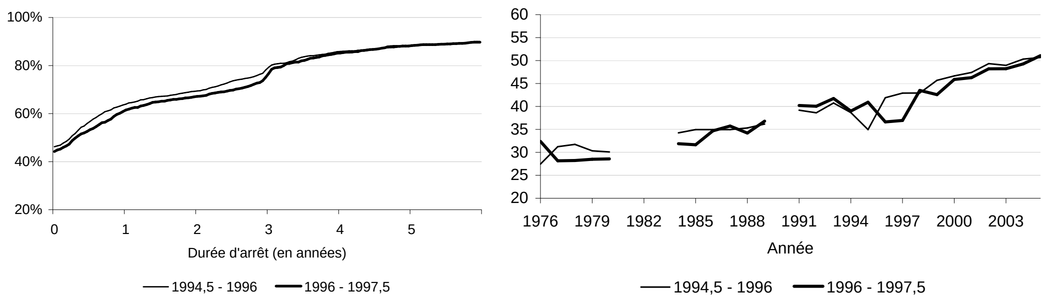
FIGURE 2.1 – Durée de l'arrêt et salaires, pour les mères d'un enfant de rang 2



(a) Avant réforme vs. avant réforme : Enfant né pendant [1991.5, 1993] vs. [1993, 1994.5]



(b) Avant réforme vs. après réforme : Enfant né pendant [1993, 1994.5] vs. [1994.5, 1996]



(c) Après réforme vs. après réforme : Enfant né pendant [1994.5, 1996] vs. [1996, 1997.5]

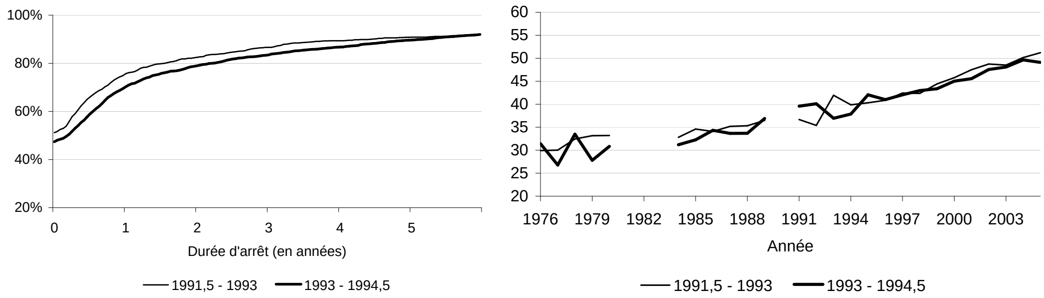
Notes : La figure 2.1 porte sur les femmes ayant eu un enfant de rang 2 entre juillet 1991 et juin 1997. Les graphiques de gauche représentent la fréquence cumulée des durées entre la naissance de rang 2 et le retour à l'emploi qui suit cette naissance. Les graphiques de droite reportent le salaire journalier moyen en €2005 pour ces mêmes personnes.
Lecture : Sur le graphique 2.1(b) de gauche, 80% (resp. 68%) des femmes ayant eu un enfant de rang 2 entre janvier 1993 et juin 1994 (resp. entre juillet 1994 et décembre 1995) sont en emploi deux ans après la naissance. Sur le graphique 2.1(b) de droite, les femmes ayant eu un enfant de rang 2 entre janvier 1993 et juin 1994 (resp. entre juillet 1994 et décembre 1995) et qui travaillent en 2003 ont un salaire journalier de 54 € (resp. 50 €) en 2003.

Utilisant d'autres sources de données, ce document confirme leurs résultats sur le taux d'emploi. Les Déclarations Annuelles de Données Sociales (DADS) contiennent des informations sur les différents épisodes d'emploi des travailleurs du secteur privé depuis 1976 et des fonctions publiques territoriale et hospitalière depuis les années 1980 (salaire, temps de travail, catégorie socioprofessionnelle, etc.). L'Échantillon Démographique Permanent (EDP) est un extrait des recensements depuis 1968, auquel les bulletins d'état civil ont été ajoutés. L'EDP permet entre autres de connaître les dates de naissance des enfants des personnes dont les évolutions de carrière sont étudiées. Les résultats présentés dans cette étude sont produits en exploitant le fichier apparié DADS/EDP (se reporter au chapitre 3 pour plus de détails sur l'appariement de ces deux fichiers). Les deux principaux avantages de cette source de données sur celles utilisées par Piketty (2005) et Pailhé and Solaz (2006) sont la structure longitudinale sur longue période (30 ans), avec la présence des salaires et non uniquement des durées en emploi.

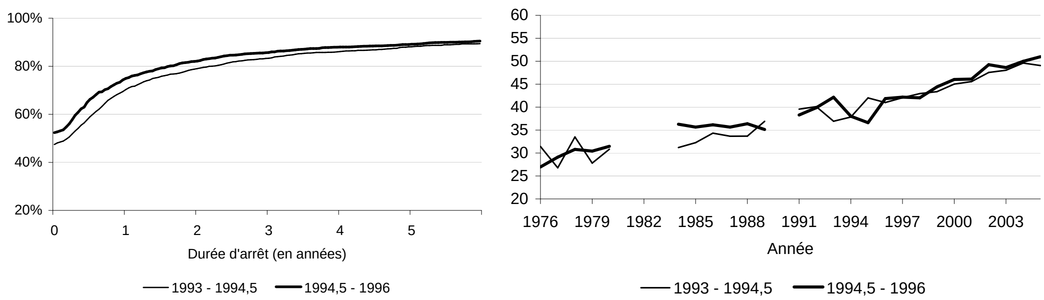
Les trois graphiques de gauche de la figure 2.1 représentent la fréquence cumulée des retours à l'emploi après la naissance d'un enfant de rang 2. La durée d'interruption est égale à la durée séparant la naissance de l'enfant du début du premier épisode d'emploi postérieur à la naissance. Différentes courbes sont tracées suivant la date de naissance de l'enfant. Le graphique du haut compare les durées d'arrêt pour des enfants nés entre juillet 1991 et décembre 1992 d'une part, et entre janvier 1993 et juin 1994 d'autre part. Toutes ces naissances ont eu lieu avant la réforme de l'APE de juillet 1994, et les parents ont donc fait face aux mêmes incitations pour se retirer du marché du travail. Les deux courbes sont très proches l'une de l'autre, comme attendu. Le graphique du milieu est lui centré sur les naissances ayant eu lieu juste avant et juste après la réforme. Cette fois-ci, les deux courbes diffèrent nettement : après la réforme, les retours à l'emploi sont moins fréquents dans les trois premières années suivant la naissance, période maximale durant laquelle on peut bénéficier de l'APE. A partir du troisième anniversaire de l'enfant, les courbes se rejoignent. Cela correspond exactement à l'effet théorique que l'APE devrait avoir sur la distribution des durées d'arrêt.

A ce stade, il faut rester prudent dans l'interprétation de ces graphiques. L'APE

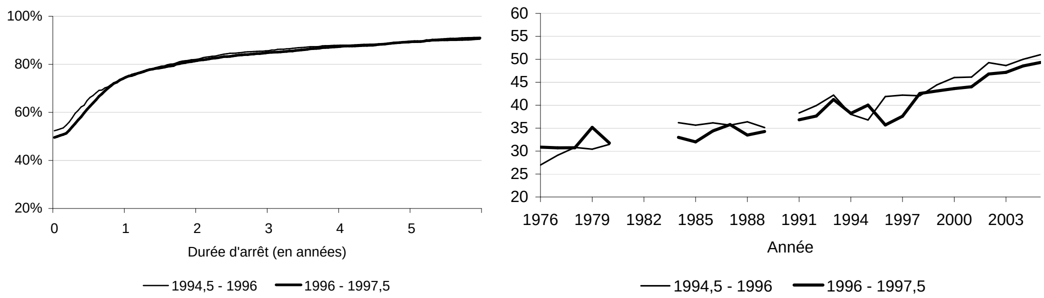
FIGURE 2.2 – Durée de l'arrêt et salaires, pour les mères d'un enfant de rang 1



(a) Avant réforme vs. avant réforme : Enfant né pendant [1991.5, 1993] vs. [1993, 1994.5]



(b) Avant réforme vs. après réforme : Enfant né pendant [1993, 1994.5] vs. [1994.5, 1996]



(c) Après réforme vs. après réforme : Enfant né pendant [1994.5, 1996] vs. [1996, 1997.5]

Notes : La figure 2.2 porte sur les femmes ayant eu un enfant de rang 1 entre juillet 1991 et juin 1997. Les graphiques de gauche représentent la fréquence cumulée des durées entre la naissance de rang 1 et le retour à l'emploi qui suit cette naissance. Les graphiques de droite reportent le salaire journalier moyen en €2005 pour ces mêmes personnes.
Lecture : se reporter à la Figure 2.1.

est certes un excellent candidat pour expliquer le changement de comportement que l'on constate entre ces deux graphiques, mais il est impossible d'affirmer avec certitude qu'il s'agit de l'unique explication. Il se pourrait que d'autres événements ayant eu lieu à la même période aient incité les mères à allonger la durée de leur arrêt (une conjoncture économique qui se dégrade, avec un chômage élevé, pourrait retarder le retour à l'emploi). Si ces événements ont aussi une influence directe sur le niveau de salaire, il ne sera pas possible d'attribuer de manière causale d'éventuelles différences de salaire à la variation de la durée d'arrêt.

Deux éléments viennent cependant conforter l'hypothèse que l'APE explique la majeure partie des différences entre ces deux graphiques. Le graphique 2.1(c) montre que la distribution des durées est restée stable entre juillet 1994 et juillet 1997, alors que les conditions macroéconomiques commençaient à s'améliorer à la fin de cette période[2]. La durée d'arrêt semble donc être peu sensible au contexte économique. De plus, aucun changement notable de durée n'est à signaler sur la période 1991-1997 pour les naissances de rang 1 (figure 2.2). Les parents d'un enfant de rang 1, non concernés par la réforme de l'APE en 1994, ont pourtant été confrontés aux mêmes variations de contexte économique que les parents de deux enfants.

Finalement, il paraît raisonnable de dire que la réforme de l'APE en 1994 a bien modifié la durée des retraits du marché du travail pour les personnes nouvellement éligibles, et que les différences de durées observées à partir de 1994 sont principalement dues à cette réforme. Par ailleurs, il n'y a pas de mécanisme évident indiquant que cette réforme pourrait avoir un effet direct sur les salaires. Elle semble donc être un instrument valide pour mesurer l'impact causal de la durée d'interruption sur les salaires.

[2]La déformation de la courbe des durées observée entre juillet 1994 et décembre 1995 semble même s'amplifier très légèrement. Cela est certainement dû au fait que l'ensemble de la population était mieux informé en 1996 qu'en 1994 de l'existence de la réforme de l'APE.

2.2 Comment mesurer l'impact de la durée de l'arrêt sur les salaires ?

2.2.1 Plusieurs phénomènes se superposent

Une méthodologie naïve consisterait à comparer le salaire des femmes qui ont repris le travail dès la fin du congé maternité, avec le salaire de celles qui se sont arrêtées, disons, un an. Si l'on observait un écart salarial entre ces deux catégories de femmes, il serait alors tentant de l'interpréter comme étant causé par l'année d'interruption de carrière.

Bien qu'attrayante, car extrêmement simple, cette approche est cependant incorrecte. Il est en effet probable que la différence de salaire observée soit due, en partie ou en totalité, à d'autres facteurs que la durée d'arrêt. Plusieurs mécanismes peuvent ainsi expliquer qu'un différentiel salarial existe, sans que la durée de l'arrêt en soit la cause. Ces mécanismes sont basés sur le fait que le choix de la durée d'arrêt est lié à des caractéristiques individuelles qui ont elles-mêmes une influence sur les évolutions de salaire. Par exemple, il est probable que les femmes qui limitent au maximum la durée de leur interruption soient en moyenne celles qui accordent le plus d'importance à l'évolution de leur carrière professionnelle. Elles prendraient de ce fait un congé plus court que celles qui privilégient leur vie familiale. Cela pourrait contribuer à créer une différence de salaire entre les femmes avec une durée d'arrêt courte et celles avec une durée plus élevée. De même, les salariées qui ont les perspectives de carrières les moins favorables (moins diplômées, environnement de travail moins stimulant, faibles perspectives de promotion, etc.) sont peut-être celles qui interrompent leur carrière le plus longtemps[3]. Là encore, la différence de salaire observée après le retour à l'emploi serait alors le reflet de caractéristiques des salariées et de leur emploi, plutôt que l'effet de la durée de l'arrêt.

Compte tenu de ces différents éléments, il n'est en général pas possible de me-

[3] Piketty (2005) constate ainsi que les femmes peu diplômées bénéficient plus fréquemment de l'APE et qu'elles ont une durée d'arrêt plus élevée que les femmes diplômées. Une explication possible est que le montant de l'allocation est forfaitaire, et donc que l'incitation financière est relativement plus importante pour les salariées qui ont un faible salaire.

surer directement l'impact de la durée de l'arrêt sur les salaires. Si l'on parvient à observer l'effet cumulé dû à l'ensemble des mécanismes en jeu, on ne peut habituellement pas distinguer l'impact propre à la durée de l'effet des autres mécanismes cités ci-dessus.

2.2.2 Les différences de différences

Une manière de procéder pour identifier l'impact causal de la durée d'interruption sur les salaires est de construire un modèle structurel, qui préciserait le lien entre présence sur le marché du travail, durée d'interruption et salaire. Celui-ci prendrait en compte l'ensemble des interactions entre les caractéristiques observées et inobservées des individus, celles de leur emploi, de leur employeur et de la conjoncture économique, ceci afin de modéliser conjointement les choix d'interruption d'activité et les salaires. L'estimation d'un tel modèle nécessite des hypothèses sur la spécification des équations (forme fonctionnelle, loi des termes d'erreur, etc.), la sélection endogène des personnes qui sont présentes sur le marché du travail, etc. Au final, les résultats produits par un tel modèle sont conditionnés par la spécification retenue.

Certaines circonstances particulières permettent toutefois d'isoler l'effet de la durée de l'interruption sur les salaires, tout en s'affranchissant des contraintes de l'estimation d'un modèle structurel. Un tel cas favorable se produit quand l'on dispose d'une réforme satisfaisant les deux critères suivants. D'une part, cette réforme doit avoir modifié le choix de la durée de l'arrêt pour un grand nombre de salariées. La partie 2.1.2 montre que l'extension de l'APE en 1994 a effectivement modifié la distribution des durées d'interruption de carrière. D'autre part elle ne doit pas avoir eu d'influence sur les autres variables individuelles ayant un impact direct sur le salaire (être plus ou moins fortement carriériste, avoir de bonnes perspectives de salaire, etc.). Autrement dit, l'unique manière pour ce changement législatif d'affecter les salaires est d'activer un mécanisme causal indirect passant par la durée de l'arrêt : il modifie la durée de l'arrêt, et ce changement de durée a un impact sur le salaire.

Ainsi, si l'on constate une différence de salaire entre les femmes selon qu'elles ont eu leur enfant sous l'ancien ou le nouveau régime législatif, ce différentiel pourra

TABLE 2.1 – Statistiques descriptives

	Mères d'un enfant de rang 2 né dans			
	[1991.5, 1993]	[1993, 1994.5]	[1994.5, 1996]	[1996, 1997.5]
Année de naissance	1963.3	1964.1	1965.6	1966.5
Age à la première naissance	24.8	25.4	25.4	25.9
Age à la deuxième naissance	29.0	29.7	29.7	30.2
Nombre d'enfants				
2	59%	63%	66%	69%
3	31%	30%	28%	26%
4 ou plus	10%	7%	6%	5%
Nombre d'obs.	932	1065	1153	1202

	Mères d'un enfant de rang 1 né dans			
	[1991.5, 1993]	[1993, 1994.5]	[1994.5, 1996]	[1996, 1997.5]
Année de naissance	1965.9	1966.9	1968.0	1969.4
Age à la première naissance	26.4	26.8	27.2	27.3
Nombre d'enfants				
1	21%	22%	25%	24%
2	53%	53%	51%	54%
3	21%	21%	20%	19%
4 ou plus	5%	4%	4%	3%
Nombre d'obs.	1467	1601	1449	1606

Note : Le nombre d'enfants correspond au nombre d'enfants nés avant le 31 décembre 2005.

être attribué à la variation de la durée du retrait du marché du travail. En revanche, si le niveau de salaire ne change pas avec l'application de la réforme alors que la durée a été modifiée, cela signifie forcément que la durée de l'arrêt n'a pas d'effet sur les salaires. Cette méthode permet donc de savoir si cet impact causal existe réellement, et dans l'affirmative, de le quantifier.

L'hypothèse implicite pour appliquer une telle méthode est que l'on dispose d'une population témoin qui "ressemble" en tous points aux femmes concernées par la réforme (âge, éducation, catégories socio-professionnelles, proportion de femmes carriéristes, situation maritale, etc.). La seule différence entre ces deux sous-populations est que l'une a été affectée par la réforme, et l'autre non. Puisque ces deux sous-populations sont similaires, les relations structurelles définissant la durée d'interruption et le niveau de salaires doivent être les mêmes sur ces deux groupes de personnes. Le recours à un groupe témoin permet ainsi de quantifier l'impact des variations de variables observables et inobservables sur les salaires dans un monde où la réforme n'aurait pas eu lieu. Un choix naturel pour un tel groupe témoin est la sous-population qui aurait été affectée par la réforme si celle-ci avait été mise en place un peu plus tôt. Ainsi, les femmes ayant eu un enfant un an avant l'application de la réforme et qui auraient été éligibles si la réforme avait été implémentée un an plus tôt constituent un groupe de contrôle particulièrement intéressant : les caractéristiques de cette sous-population sont par définition extrêmement proches de celles des personnes affectées par la réforme, vu que la seule dimension qui les différencie théoriquement est la date de naissance de leur enfant[4]. Le tableau 2.1 confirme que les caractéristiques de ces deux groupes sont effectivement très similaires.

L'impact causal de la durée d'interruption sur les salaires est alors estimé de la manière suivante : on calcule la différence de salaire avant et après la naissance pour la sous-population affectée par la réforme, et on lui ôte la même différence de salaire mesurée sur la population témoin. La grande qualité de cette stratégie en *différences de différences* est de permettre l'identification tout en s'affranchissant des contraintes inhérentes à l'estimation d'un modèle structurel. Il n'est en effet pas nécessaire de connaître la manière dont toutes les variables observables (âge, éducation, nombre d'enfant, situation maritale, revenu du conjoint, etc.) et inobservables (motivation, talent, carriérisme, environnement de travail stimulant, etc.) interagissent dans la formation des choix d'interruption de carrière et des salaires.

[4]La loi a été votée le 25 juillet 1994, et la réforme n'a pas été annoncée longtemps à l'avance, ce qui exclut tout effet d'anticipation de la part des futurs parents qui auraient pu vouloir retarder la naissance de leur deuxième enfant afin de devenir éligible. Il se pourrait également que la réforme ait eu un impact positif sur la fécondité, en incitant certaines femmes à avoir un deuxième enfant. Piketty (2005) estime que cet effet incitatif sur la fécondité de rang 2 a été faible.

Il suffit uniquement que les véritables équations structurelles (inconnues) guidant les choix de carrière des femmes restent les mêmes sur une courte période encadrant la date de la réforme. Cette hypothèse de *continuité* est nettement moins lourde que celles présentes dans un modèle structurel, et elle semble raisonnable dans ce contexte. Il en découle que les résultats d'une telle estimation sont plus crédibles et plus robustes, parce que moins soumis à une éventuelle mauvaise spécification.

2.2.3 Les triples différences

Ce raisonnement en doubles différences est valable à condition que les deux groupes aient subi les mêmes variations d'environnement dans l'intervalle de temps où les salaires sont mesurés de part et d'autre de la naissance. Cela autorise donc par exemple la conjoncture économique à se dégrader brusquement au moment où la réforme est appliquée : l'observation de la différence de salaires entre avant et après la naissance sur le groupe témoin capture cet effet, tandis que la variation salariale sur le groupe affecté par la réforme mesure la somme des effets créés par les variations de durées d'interruption et de conjoncture économique. L'estimateur en doubles différences isole donc bien l'impact causal recherché dans un tel cas. En revanche, si les femmes ayant eu un enfant avant juillet 1994 sont affectées différemment par la conjoncture de celles ayant eu leur enfant après cette date, l'estimation en doubles différences va être biaisée. L'identification est cependant encore possible, à condition de disposer d'un estimateur de cet éventuel biais. Cela peut par exemple passer par une hypothèse sur les femmes ayant eu un enfant de rang 1 au moment où la réforme était appliquée aux parents d'un enfant de rang 2. En supposant que ce groupe de personnes a subi ce même éventuel impact différencié de la conjoncture en fonction de la date de naissance de l'enfant, il permet d'en estimer l'ampleur, pour ensuite corriger l'estimateur en doubles différences[5]. On parle alors d'estimation en *triples différences*.

Il faut également que les caractéristiques des femmes ayant un enfant de rang

[5]Les mères d'enfants de rang 3 constituent aussi un groupe de contrôle potentiel. Les données ne contiennent cependant pas suffisamment de naissances de rang 3 pour obtenir des résultats significatifs avec un tel groupe de contrôle.

2 avant et après la réforme soient les mêmes. Or le tableau 2.1 indique que l'âge de la mère à la naissance du deuxième enfant semble augmenter au fil des générations, dans des proportions comparables à celles constatées par Pison (2009). Si certaines caractéristiques changent dans le temps (plus diplômées, plus âgées, plus carriéristes, etc.), la progression salariale observée agrégera les effets de la durée d'interruption et de l'évolution de ces caractéristiques. Il n'est pas clair que ces effets de composition de population se compenseront dans la double différence, et l'estimateur en différences de différences sera donc a priori biaisé.

De plus, ces changements progressifs de caractéristiques chez les mères pourraient créer une tendance temporelle dans les choix de durée d'interruption tels qu'ils sont observés. Une telle tendance sur les durées d'interruption pourrait également exister si le comportement des mères évolue vers une présence plus importante sur le marché du travail. La différence de durée entre les deux populations représentera alors l'impact de la réforme augmenté de la variation tendancielle de durée. Par suite, cela pourrait biaiser l'estimation en différences de différences de l'impact salarial normalisé pour une année d'interruption.

Dans les deux cas, il est cependant probable que ces biais potentiels soient ici de faible ampleur : les deux populations comparées sont deux cohortes distantes de seulement 18 mois, alors que les tendances ne sont observées qu'à l'échelle de la décennie sur les taux d'emploi (se reporter au chapitre 3 ou à Piketty 2005) et les comportements liés à la fécondité comme l'âge moyen à la maternité (Pison 2009). Si l'on veut néanmoins tenir compte de cet éventuel phénomène, il faut disposer d'un groupe d'individus permettant de mesurer ces tendances. Là encore, cela implique des hypothèses supplémentaires, pour justifier le choix du groupe de contrôle utilisé pour les triples différences. Les femmes ayant eu un enfant de rang 2 deux ou trois ans de part et d'autre de la réforme peuvent ainsi permettre de contrôler d'éventuels changements tendanciels propres aux mères de rang 2 (composition de population, choix de la durée d'interruption, opportunités sur le marché du travail, etc.). Il est aussi possible d'utiliser les femmes ayant eu un enfant de rang 1 au moment où la réforme était étendue aux parents d'un enfant de rang 2 pour tenir compte des variations temporelles de la durée d'interruption.

Le choix du groupe de contrôle est primordial pour la qualité des estimations, que ce soit pour les doubles ou pour les triples différences. Dans le premier cas, un groupe naturel se dégage : les femmes ayant eu un enfant de rang 2 juste avant l'application de la réforme. Dans le second cas, plusieurs groupes sont a priori disponibles, chacun étant valide sous un jeu d'hypothèses différent. La partie 2.3 précise les conditions de validité de chacun des groupes de contrôle, et présente les résultats des estimations.

2.3 L'impact de la durée de l'arrêt sur les salaires

2.3.1 Résultats

Les graphiques de droite de la figure 2.1 reportent les évolutions de salaire des femmes qui ont eu un deuxième enfant. Le salaire moyen une année civile donnée est calculé sur l'ensemble des femmes qui ont travaillé cette année là. Comme précédemment, chaque courbe correspond à l'ensemble des femmes qui ont eu ce deuxième enfant durant une période donnée. Alors que leurs salaires étaient relativement proches jusqu'à l'année de la naissance, un écart semble se créer suite à la naissance entre les femmes dont l'enfant est né juste après la réforme de l'APE et celles dont l'enfant est né juste avant la réforme (graphique 2.1(b)). En revanche, ce phénomène ne s'observe ni pour les naissances avant la réforme (graphique 2.1(a)), ni celles après la réforme (graphique 2.1(c)).

Dans le tableau 2.2, nous nous intéressons à la progression du salaire moyen entre avant la naissance et une fois que la majorité des femmes est revenue en emploi (voir l'annexe pour la manière dont sont calculés ces salaires moyens). Cette progression est de 29.4% si l'enfant de rang 2 est né entre janvier 1993 et juin 1994, contre seulement 26.5% pour les femmes ayant eu leur deuxième enfant entre juillet 1994 et décembre 1995. Ces dernières ont donc connu une progression salariale plus faible de 2.8%, significative au seuil de 5%. Cette progression moindre est la conséquence de l'allongement de la durée d'interruption de carrière créé par la réforme de l'APE. Sachant que la durée moyenne d'interruption a augmenté de 0.27 année à l'occasion de cette réforme, cela correspond à une baisse moyenne de salaire d'environ 10%

Table 2.2 – Variations de salaire et de durée

	Salaire avant la naissance (1)	Salaire après la naissance (2)	Variation avant/après $\frac{(2)-(1)}{(1)}$	Durée d'interruption
Rang 1				
[1993, 1994.5]	37.98***	47.62***	0.254***	1.24***
	(0.19)	(0.34)	(0.0075)	(0.023)
[1994.5, 1996]	39.51***	49.09***	0.242***	1.07***
	(0.24)	(0.34)	(0.0078)	(0.026)
Différence			-0.012	-0.17***
			(0.011)	(0.035)
Rang 2				
[1991.5, 1993]	38.76***	48.11***	0.241***	1.19***
	(0.27)	(0.46)	(.0084)	(0.033)
[1993, 1994.5]	39.93***	51.65***	0.294***	1.18***
	(0.28)	(0.46)	(0.0105)	(0.028)
Différence			0.055***	-0.005
			(0.013)	(0.044)
Rang 2				
[1993, 1994.5]	39.93***	51.65***	0.294***	1.18***
	(0.28)	(0.44)	(0.0105)	(0.028)
[1994.5, 1996]	38.90***	49.23***	0.265***	1.45***
	(0.26)	(0.38)	(0.0084)	(0.028)
Différence			-0.028**	0.27***
			(0.013)	(0.040)
Rang 2				
[1994.5, 1996]	38.90***	49.23***	0.265***	1.45***
	(0.26)	(0.38)	(0.0084)	(0.028)
[1996, 1997.5]	39.88***	48.92***	0.227***	1.54***
	(0.24)	(0.38)	(0.0075)	(0.031)
Différence			-0.039***	0.09**
			(0.011)	(0.046)

Notes : Salaires journaliers moyens en €2005, durées d'interruption en années. Ecart-types entre parenthèses, obtenus par bootstrap. *, **, et *** signifient que les coefficients sont significatifs aux seuils de 10%, 5% et 1% respectivement. Voire l'annexe pour la méthodologie concernant le calcul des salaires moyens avant et après la naissance.

par année d'interruption de carrière (tableau 2.3). Il faut interpréter ce chiffre de la manière suivante : interrompre sa carrière une année supplémentaire diminue le salaire journalier moyen de 10% une fois le retour à l'emploi effectué. Cet effet est statistiquement significatif au seuil de 5%. D'après le graphique 2.1(b), il semble que cette diminution intervienne très rapidement après le retour à l'emploi. Son ampleur globalement constante dans les 10 ans après la naissance indique qu'il n'y aurait pas de rattrapage progressif après la baisse initiale. Il ne semble pas non plus s'agir d'une reprise d'emploi à un salaire "normal" suivie d'une stagnation, ou d'une progression salariale moins forte pour les femmes qui se sont interrompues plus longtemps.

Un calcul similaire peut être mené sur les femmes ayant eu un enfant de rang 1 durant la période janvier 1993 - décembre 1995. Elles ont en effet vécu les mêmes variations d'environnement économique que les mères d'un enfant de rang 2, et elles n'ont pas été concernées par la réforme de l'APE en 1994. Sous l'hypothèse que les mères d'enfants de rangs 1 et 2 ont été affectées de la même manière par la conjoncture, ce groupe "placebo" nous permet donc de tester si des changements macroéconomiques contemporains de la réforme de l'APE auraient affecté différemment les femmes ayant un enfant avant ou après juillet 1994. Un tel impact différencié pourrait en effet être à l'origine d'une part de l'écart de salaire observé entre les mères d'un enfant de rang 2, selon que la naissance a eu lieu avant ou après juillet 1994. Le tableau 2.2 montre que les salaires des femmes ayant eu leur premier enfant entre juillet 1994 et décembre 1995 ont progressé très légèrement moins vite que ceux des femmes ayant eu leur premier enfant entre janvier 1993 et juin 1994 (-1.2%). Cet écart de salaire n'est pas significatif au seuil de 10%, ce qui indique que les variations d'environnement économique ne semblent pas être à l'origine du différentiel de salaire observé pour les mères d'un enfant de rang 2.

Sous des hypothèses différentes, ce même groupe des femmes ayant eu un enfant de rang 1 entre 1993 et 1995 permet d'obtenir un autre estimateur de la hausse de durée d'interruption provoquée par la réforme de l'APE. Si les décisions d'interruption de carrière suite à une naissance évoluent de manière similaire pour les naissances de rang 1 et de rang 2 en l'absence de réforme entre 1993 et 1995, elles peuvent servir à estimer la variation de durée que l'on aurait observée entre 1993 et 1995 pour les naissances de rang 2 si la réforme de l'APE n'avait pas existé. Utiliser les mères d'enfant de rang 1 comme groupe de contrôle revient à supposer qu'à la place de la hausse de la durée observée en juillet 1994 pour les mères de rang 2, il y aurait eu la même baisse que celle observée sur le graphique 2.2(b) pour les mères de rang 1. Cela conduit alors à former un estimateur en doubles différences pour la durée d'interruption. Les salaires peuvent être corrigés de manière similaire, en déduisant des variations de salaire des mères de rang 2 les variations de salaire observées pour les mères de rang 1. Cet estimateur en triples différences permet de prendre en compte un effet différencié de la conjoncture sur les salaires, selon que les femmes ont eu un enfant avant ou après la réforme. Il repose sur l'hypothèse que les variations de conjoncture économique ont affecté de la même manière les salaires des mères d'un enfant de rang 1 et de rang 2. Les variations de salaire corrigées des effets macroéconomiques "différenciés" pour les mères d'un enfant de rang 2 sont de 1.7% (au lieu de 2.8% sans en tenir compte), tandis que le différentiel de durée d'interruption devient 0.44 année (au lieu de 0.27). L'impact causal d'une année d'interruption sur les salaires est alors ramené à une diminution de salaire de 3.8%, non significative au seuil de 10% (tableau 2.3).

Deux autres groupes de contrôle pour les triples différences sont également disponibles : les femmes ayant eu un enfant de rang 2 entre juillet 1991 et juin 1994, et entre juillet 1994 et juin 1997. Les femmes de ces groupes ont toutes eu un deuxième enfant, ce qui les rend plus proches du groupe d'intérêt que le groupe de contrôle du paragraphe précédent composé de naissances de premier rang. Elles permettent donc a priori de mieux estimer la tendance dans les choix de durée d'interruption. Cette hypothèse semble être réaliste avec le premier de ces groupes. Notons toutefois qu'elle pourrait ne pas être valide pour le second groupe si les changements de comportements induits par la réforme ont été progressifs dans les années suivant 1994. Dans ce cas, la réforme n'aurait pas eu un effet uniforme sur les choix de durée d'interruption au sein des femmes ayant un enfant de rang 2 entre juillet 1994 et juin 1997. La note 2 indique que cette hypothèse ne peut pas être écartée.

Concernant la qualité de ces groupes de contrôle pour mesurer les variations

TABLE 2.3 – Impact causal d'une année d'interruption sur les salaires

Groupe de contrôle des doubles différences	Groupe de contrôle des triples différences	Rendement d'une année d'interruption
Mères d'un enfant de rang 2 dans [1993, 1994.5]		-0.105** (0.051)
Mères d'un enfant de rang 2 dans [1993, 1994.5]	Mères d'un enfant de rang 1 dans [1993, 1996]	-0.038 (0.042)
Mères d'un enfant de rang 2 dans [1993, 1994.5]	Mères d'un enfant de rang 2 dans [1991.5, 1994.5]	-0.302*** (0.10)
Mères d'un enfant de rang 2 dans [1993, 1994.5]	Mères d'un enfant de rang 2 dans [1994.5, 1997.5]	0.060 (0.14)

Notes : Ecart-types entre parenthèses, obtenus par bootstrap avec 200 itérations. *, **, et *** signifient que les coefficients sont significatifs aux seuils de 10%, 5% et 1% respectivement. Voire l'annexe pour le détail des calculs des salaires moyens avant et après la naissance.

salariales, disposer de mères d'enfants de rang 2 et non de rang 1 semble être là aussi un avantage. En revanche, les conditions macroéconomiques au moment de la naissance de l'enfant des femmes de ces deux groupes étaient différentes de celles existant pour le groupe d'intérêt, phénomène dont il était tenu compte dans le paragraphe ci-dessus. Cela nous place alors dans un cadre dual du cas précédent : il n'est plus nécessaire de supposer que les choix de durée d'interruption des mères de rang 1 et 2 évoluent de la même manière autour de 1994, mais il faut maintenant supposer que les variations d'environnement économique ont eu le même effet différencié pour des naissances entre 1993 et 1995 et entre 1991 et 1994 (resp. 1994 et 1997). Cette dernière hypothèse est certainement fragile dans notre contexte, puisque la France a connu une période de crise économique autour de 1993.

L'utilisation des mères d'enfants nés entre juillet 1991 et juin 1994 comme groupe de contrôle conduit à un impact causal de -30%, significatif au seuil de 1% (tableau 2.3). L'ampleur de ce coefficient semble confirmer que l'hypothèse portant sur la stabilité des conditions macroéconomiques entre 1991 et 1995 n'est effectivement pas valide : la brutale dégradation de l'économie après 1993 explique certainement une grande partie de cette importante baisse de salaire. Avec l'autre groupe de contrôle, l'estimation conduit à un impact positif, non significativement différent de zéro. Cette fois-ci, il est probable que l'amélioration des conditions économiques après la crise de 1993 biaise à la hausse les résultats des estimations. Il convient donc de rester prudent dans l'interprétation des résultats de ces deux dernières estimations, car ils sont fondés sur des hypothèses que la crise économique de 1993 rend très fragiles.

Bien que l'ampleur de l'effet varie avec le choix du groupe de référence, les salaires semblent donc dépendre négativement de la durée d'interruption de carrière. L'estimation en doubles différences indique qu'une année d'interruption de carrière diminue le salaire journalier d'environ 10% pour les femmes qui ont été affectées par le réforme de l'APE en 1994.

2.3.2 Commentaires

Plusieurs mécanismes peuvent être avancés pour expliquer par quels canaux transite une partie de cet impact causal de la durée de l'interruption sur les salaires. Tout d'abord, une année passée hors du marché du travail est aussi une année où la salariée n'accumule pas d'expérience professionnelle. Or l'expérience est l'un des facteurs expliquant les augmentations de salaire, le rendement d'une année d'expérience étant estimé autour de 2% sur données françaises (Bardaji et al. 2003, Koubi 2003). L'impact négatif de la durée d'interruption capture certainement cette perte de salaire due à un nombre d'année d'expérience plus faible.

De plus, chaque période passée hors du marché du travail peut provoquer une dégradation du capital humain, dégradation d'autant plus importante que la durée d'arrêt de travail est longue. Ceci n'est pas propre aux arrêts consécutifs à une naissance, et concerne la plupart des causes de retrait du marché du travail (le chômage en particulier). Cette perte de capital humain implique que le salarié a une productivité moindre lors de son retour à l'emploi, ce qui se traduit par un salaire moins élevé sur le marché du travail. Albrecht et al. (1999) estiment sur données suédoises que le capital humain des femmes se déprécie plus vite lors d'un épisode de chômage que lors d'un arrêt après une naissance.

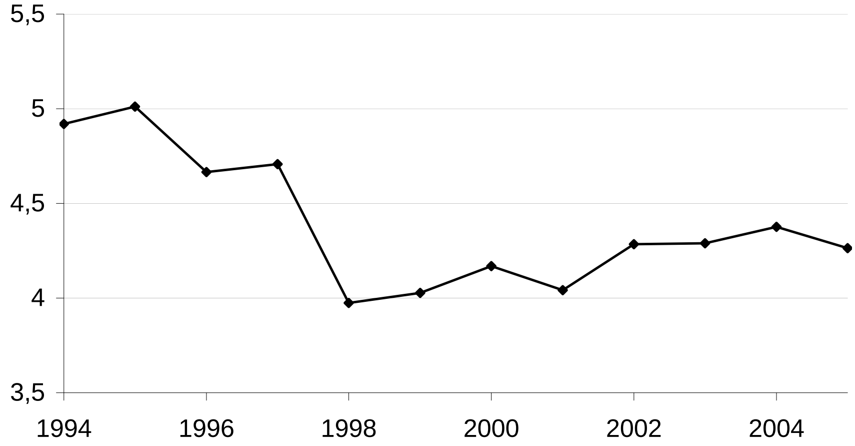
En revanche, la croissance des salaires réels durant l'interruption ne joue théoriquement aucun rôle dans les variations de salaire observées suite à l'arrêt : le Congé Parental d'Éducation prévoit que le salarié doit percevoir à son retour à l'emploi au minimum son ancien salaire plus l'augmentation salariale moyenne ayant été observée dans l'entreprise durant l'arrêt[6]. Comme par construction cet article étudie les évolutions de carrière des femmes qui travaillent avant la naissance, la plupart de celles qui sont éligibles à l'APE sont aussi éligibles au CPE.

La sélection de l'échantillon ne semble pas non plus à l'origine de la baisse de salaire. Le salaire moyen avant (resp. après) une naissance est calculé sur l'ensemble des femmes qui sont présentes dans les DADS dans les années précédant (resp. suivant) la naissance. A condition que le processus de participation au marché du travail ne change pas radicalement et brutalement en 1994 (les deux cohortes sont distantes de seulement 18 mois), l'estimation en doubles différences prend en compte le fait que toutes les mères ne sont pas en emploi durant toute la vie active. Cependant, la réforme de l'APE ayant perturbé la distribution des durées d'interruption inférieures à 3 ans (graphique 2.1(b)), les caractéristiques des femmes qui retravaillent dans les 3 premières années après la naissance ne sont certainement pas les mêmes selon que la naissance a eu lieu avant ou après la réforme. Le graphique 2.1(b) semble cependant indiquer que la réforme n'a pas d'impact sur la proportion des femmes qui retravaillent après une interruption supérieure à 3 ans. C'est pourquoi les salaires moyens après la naissances ont été calculés sur des années éloignées au minimum de 3 ans de l'année de naissance de l'enfant (voire annexe). De cette manière, les populations sur lesquelles les salaires moyens sont estimés sont comparables en termes de caractéristiques, et cela ne risque donc pas d'introduire de biais dans l'estimation de l'impact causal.

Des variations du nombre d'heures travaillées annuellement pourraient expliquer

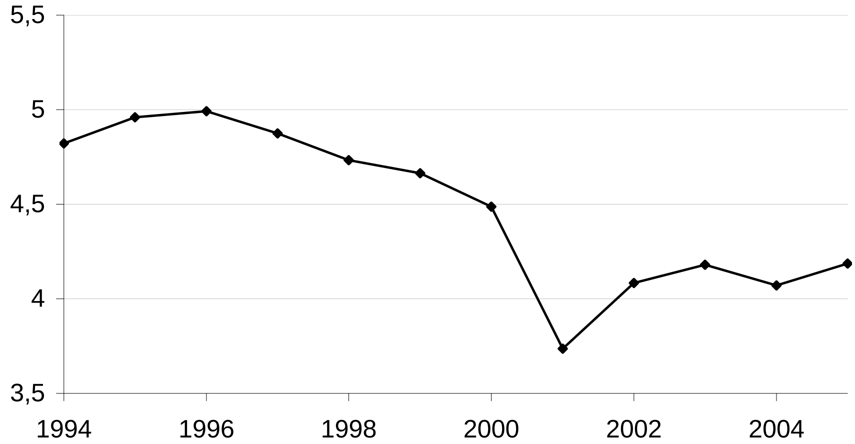
[6] En toute rigueur, la garantie de revalorisation en fonction du salaire moyen n'évite cependant pas tout décrochement par rapport aux femmes qui ne se sont pas arrêtées : la dynamique du salaire moyen inclut l'effet de renouvellement, et elle est donc moins rapide que celle des salaires des personnes en emploi.

FIGURE 2.3 – Nombre d'heures travaillées par jour, pour les naissances de rang 2 en 1998



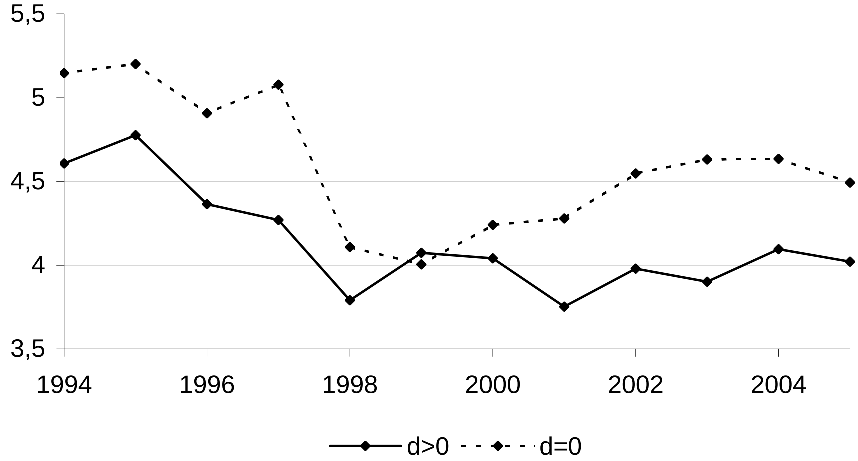
Note : Le graphique représente le nombre moyen d'heures travaillées par jour rémunéré, pour les femmes qui ont eu un enfant de rang 2 en 1998.
Lecture : Pour les femmes ayant eu un enfant de rang 2 en 1998 qui travaillaient en 2004, le nombre moyen d'heures travaillées par jour était de 4.4 heures en 2004.

FIGURE 2.4 – Nombre d'heures travaillées par jour, pour les naissances de rang 2 en 2001



Note : Le graphique représente le nombre moyen d'heures travaillées par jour rémunéré, pour les femmes qui ont eu un enfant de rang 2 en 2001.
Lecture : Pour les femmes ayant eu un enfant de rang 2 en 2001 qui travaillaient en 1996, le nombre moyen d'heures travaillées par jour était de 5.0 heures en 1996.

FIGURE 2.5 – Nombre d'heures travaillées par jour, pour les naissances de rang 2 en 1998



Note : Le graphique représente le nombre moyen d'heures travaillées par jour rémunéré, pour les femmes qui ont eu un enfant de rang 2 en 1998. La courbe en traits pointillés concerne les femmes qui ont quitté le marché du travail suite à la naissance de rang 2 uniquement durant la durée légale du congé de maternité ($d = 0$), la courbe en traits pleins concerne celles qui se sont retirées plus longtemps que le congé de maternité ($d > 0$).
Lecture : Pour les femmes ayant eu un enfant de rang 2 en 1998, qui se sont arrêtées plus longtemps que la durée légale du congé de maternité, et qui travaillaient en 2002, le nombre moyen d'heures travaillées par jour était de 4.0 heures en 2002.

la baisse de salaire observée. Le salaire tel qu'il est mesuré dans cet article est le salaire annuel divisé par la longueur en jours de la période d'emploi sur l'année[7]. Cela correspond donc à un salaire journalier. Or une baisse de salaire journalier peut avoir lieu sans que le salaire horaire ne baisse simultanément. Il suffit pour cela que le nombre d'heures travaillées par jour diminue dans les mêmes proportions que le salaire journalier. Les DADS ne contiennent l'information sur le nombre d'heures travaillées que depuis 1994, ce qui interdit de calculer un salaire horaire pour les années travaillées avant 1994. Il n'est donc pas possible de réaliser les estimations de la partie 2.2 en remplaçant le salaire journalier par le salaire horaire. Afin de quantifier l'ampleur des variations de temps de travail, le graphique 2.3 reporte

[7]Les DADS renseignent sur le jour de début et le jour de fin de la période de rémunération, et non sur le nombre de jours rémunérés proprement dit. Ainsi une personne qui travaille toute une année à temps partiel 4 jours par semaine aura une période d'emploi d'un an, et non de 4/5 d'une année.

pour les femmes qui ont eu un deuxième enfant en 1998 l'évolution du nombre moyen d'heures travaillées par jour rémunéré entre 1994 et 2005. Le graphique 2.4 représente la même chose pour une femme ayant eu un enfant de rang 2 en 2001. Ne disposant pas de données sur une période suffisamment longue pour avoir simultanément beaucoup de points de part et d'autre de l'année de naissance pour une même cohorte de naissance, ces deux années de naissance 1998 et 2001 ont été choisies pour avoir un aperçu de ce qu'il se passe “longtemps” après et avant la naissance respectivement. On y constate une diminution d'environ une demi-heure de nombre d'heures quotidiennes après la naissance, ce qui correspond à une baisse d'environ 10%. D'après le graphique 2.5, l'ordre de grandeur de cette diminution ne semble pas dépendre du fait que la durée du retrait du marché du travail soit supérieure ou non à la durée légale du congé de maternité. La figure 2.3 indique que la baisse paraît se gommer très partiellement dans les années qui suivent la naissance. Cependant en 2005, 7 ans après la naissance, elle semble être encore significative. L'absence de recul supplémentaire ne permet pas de dire si l'écart reste visible jusqu'à la fin de la carrière, ou s'il s'estompte complètement une fois que les enfants (ceux de rangs 1 et 2, et éventuellement de rangs supérieurs) ont grandi.

Cela pourrait expliquer une partie de l'effet de la durée de l'interruption sur les salaires, si les femmes qui s'arrêtent le plus longtemps sont incitées *à cause de cette durée d'arrêt importante* à réduire le nombre d'heures travaillées une fois revenues en emploi. Il n'est malheureusement pas possible de tester cette hypothèse avec les données disponibles dans les DADS. Les données ne permettent d'ailleurs pas non plus de savoir si la réduction du volume horaire est choisie ou subie par la salariée (par exemple, parce que l'employeur proposerait plus fréquemment des emplois à temps partiel aux mères qui reviennent sur le marché du travail après un long arrêt).

Une autre limite des données est que les DADS ne couvrent pas la totalité des personnes travaillant en France. En particulier, la fonction publique d'Etat et les indépendants ne sont pas dans le champ des DADS. Il y a donc une erreur de mesure (surestimation) sur la durée d'interruption pour les personnes dont le premier emploi après la naissance n'est pas dans le champ des DADS. De même, les DADS ne permetttent pas de faire la différence entre les périodes d'inactivité et les périodes de non-emploi, pouvant inclure des épisodes de chômage. Cela contribue également à surestimer la durée d'interruption pour les personnes au chômage après la naissance. Cela ne concerne cependant que peu de bénéficiaires de l'APE, qui sont souvent également bénéficiaires du CPE et ont donc l'assurance de retrouver un emploi chez leur ancien employeur. Enfin, et surtout, ces deux sources d'erreur de mesure sont a priori orthogonales à l'instrument utilisé (avoir eu son deuxième enfant avant ou après juillet 1994). Elles ne biaisent donc pas les estimations. De plus, si les erreurs de mesure n'étaient pas parfaitement indépendantes de l'instrument, ces deux mécanismes conduiraient à sousestimer la perte de salaire (en valeur absolue) provoquée par une année d'interruption. Cela renforcerait alors la portée des estimations en doubles différences, qui vont dans le sens d'un impact causal significatif malgré cet éventuel biais vers zéro.

2.4 Conclusion

Cet article étudie l'effet de la durée d'une interruption de carrière suite à une naissance sur les salaires après le retour à l'emploi. En s'appuyant sur la réforme de l'APE ayant eu lieu en 1994 pour l'identification, les estimations indiquent que cet impact causal est négatif, bien que la baisse de salaire causée par la durée de l'arrêt ne soit pas statistiquement significative dans toutes les spécifications. Une partie de cet impact transite certainement par des mécanismes que l'on sait décrire (perte d'expérience professionnelle, variation - choisie ou contrainte - du temps de travail), ce qui conduit à réduire la part inexpliquée de la diminution de salaire causée par la durée de l'arrêt. Cette perte de salaire semble continuer à exister près de 10 ans après le retour sur le marché du travail. La réforme de l'APE ayant eu lieu en 1994, nous ne disposons actuellement pas du recul nécessaire pour observer la fin de carrière des bénéficiaires. Il est donc pour le moment impossible d'estimer l'impact de long terme que pourrait avoir la durée de l'interruption sur les salaires.

Finalement l'APE est considérée depuis sa création comme une politique familiale, en permettant aux travailleurs de combiner de manière plus flexible vies familiale et professionnelle. Cet article montre que l'APE a aussi des conséquences sur le déroulement de la carrière des personnes qui en bénéficient. Les femmes qui ont été incitées par la réforme de 1994 à augmenter la durée de leur interruption de carrière (les *compliers* dans la terminologie d'Angrist et al. 1996) semblent avoir subi une perte de salaire à cause de cette augmentation de durée d'interruption.

Annexe

Cette annexe détaille la manière dont sont calculés les salaires.

Le salaire provient du panel DADS, qui contient pour chaque année civile la rémunération (salaire net, primes incluses) versée par un employeur à son salarié, et le nombre de jours travaillés correspondant. Ces deux variables permettent de calculer un salaire journalier, afin de prendre en compte que les personnes n'ont pas forcément toutes travaillé durant le même nombre de jours sur une année civile donnée. Quand un salarié a plusieurs employeurs successifs durant la même année civile, les différentes rémunérations sont additionnées, et de même pour les durées travaillées. Si les périodes d'emploi se chevauchent, les jours de la période de chevauchement ne sont comptés qu'une seule fois dans la durée travaillée totale.

Pour les figures 2.1 et 2.2, les salaires moyens une année donnée sont calculés sur les femmes qui sont présentes dans les DADS cette année là. En particulier, le salaire moyen pour une année postérieure à une naissance est calculé uniquement sur les femmes qui ont déjà repris un emploi.

Pour les estimations du tableau 2.2, des salaires moyens avant la naissance et après la naissance sont calculés pour chaque groupe de femmes ayant eu un enfant durant une plage donnée de 18 mois (colonnes (1) et (2) du tableau 2.2). Le salaire moyen avant la naissance consiste à faire la moyenne des salaires observés sur les 6 années précédant la naissance. Le salaire moyen après la naissance est calculé sur 6 années, une fois que 3 ans se sont écoulés depuis la naissance et que la majorité des mères est revenue sur le marché du travail. Ce délai de 3 ans correspond également à la durée maximale durant laquelle il est possible de percevoir l'APE. Passés ces 3 ans, les femmes ne sont plus éligibles à l'APE, et sont donc en théorie revenues en emploi dans les mêmes proportions que le groupe témoin non concerné par l'APE. Piketty (2005) et Pailhé and Solaz (2006) ont montré que les taux d'emploi 3 ans après la naissance sont similaires pour les mères d'un deuxième enfant né avant ou après la réforme de l'APE en 1994.

Pour les mères d'un enfant né entre juillet 1991 et décembre 1992, ces salaires moyens sont calculés respectivement sur les périodes 1984-1991[8] et 1997-2003. Pour les femmes ayant eu un enfant entre janvier 1993 et juin 1994, ces périodes sont 1985-1992 et 1998-2004. Pour les naissances comprises entre juillet 1994 et décembre 1995, ces périodes sont 1986-1993 et 1999-2005. Enfin pour les naissances entre janvier 1996 et juillet 1997, ces périodes sont 1987-1994 et 2000-2005. Les résultats des estimations sont peu sensibles à la définition de ces périodes (nombre d'années sur lesquelles les moyennes sont calculées, décalage d'une ou deux années des périodes).

[8] Les salaires ne sont pas disponibles dans le panel DADS pour l'année 1990. La moyenne des salaires sur la période 1984-1991 est donc en fait la moyenne sur la réunion des deux sous-périodes 1984-1989 et 1991.

Chapter 3

The impact of parental leave duration on later career

Sommaire

3.1 Introduction

The causal impact of periods spent out of the labor market on later wages has been widely studied. Most papers focus on unemployment spells, and conclude that unemployment duration has a negative causal impact on wages. Returns to experience are usually put forward to explain this result: wages don't grow during unemployment because no work experience is accumulated. Human capital may also deteriorate while not working, which would lead to a lower wage afterwards. Moreover potential employers might interpret the existence and the length of those unemployment spells negatively, and thus they might lower their wage offers. In the same vein, unemployed individuals might have a weaker bargaining power than employed people when they negotiate their wage during a job interview.

Apart from unemployment, several reasons can account for periods spent out of the work force. Using the Vietnam draft lottery, Angrist (1990) finds that being a Vietnam veteran lowers civilian earnings of white males. Employees may also quit their job, or find an agreement of temporary leave with their employer. In particular, parents can take a maternity or paternity leave after the birth of a child. Parental leave duration and unemployment are likely to have a different impact on subsequent career. Potential employers may interpret voluntarily withdrawing and being unemployed differently. They may also fear that parents of a young child might be less involved in their professional activities. This paper focuses on those parental leaves, and investigate whether later career is affected by such temporary withdrawals.

A whole strand of the literature is devoted to the analysis of women employment and fertility decisions. Variations in fertility and participation rates among countries would be linked to cultural (Fernandez and Fogli 2006) and religious (Berman et al. 2006) differences betwen countries. Apart from these non economic determinants, the institutional context would matter: family allowances, parental leave legislation, childcare availability and affordability would affect fertility and participation decisions around childbearing (see Del Boca and Wetzels 2008, for a comprehensive survey). Regarding parental leave laws, two types of policies are usually implemented: benefits can be provided to eligible parents who want to reduce their labor supply, and/or guarantees that they will have a job once the leave is over. The first one is aimed at parents who would be deterred from raising their child by themselves because of the implied wage loss. The second one is targeted on parents who anticipate that they will have difficulties to find a new job if they quit their current employer and temporarily stop working. These two kinds of programs are expected to increase fertility rates, and to allow more young parents to reduce their labor supply after the birth of a child. Different approaches have been used to estimate how women respond to these incentives. Francesconi (2002) and Keane and Wolpin (2002*a*, 2002*b*) build a dynamic model of participation and fertility choices. Laroque and Salanié (2008) estimate a joint structural model of participation and fertility decisions on French data, and find that fertility is sensitive to financial incentives for birth of first and third ranks[1].

To our knowledge, almost all articles on parental leave and career studied the return to work. As far as the APE is concerned, the return to work is theoretically guaranteed by the law for most beneficiaries. Piketty (2005) noted that participation rates four years after the birth are similar to those observed right before the birth. Hence it seems that mothers willing to work again don't have much trouble finding a new job. However, there are still many things to learn on the conditions under which this return to work happens. It may be possible that women have a less interesting job than the one they had before the birth. This could result in more frequent resignations. The simple fact of taking a parental leave may also scar them, and thus affect their subsequent wage growth rate (Buding and England 2001). Most of these papers focused on return to work, and did not estimate the causal impact of parental leave on later career path (wages, upward mobility/promotion, change of employer). Some of them were published too soon after the policy took place to have information on subsequent career characteristics. Others lacked information

[1]Laroque and Salanié (2008) use financial incentives provided by features of the French tax and benefit system, instead of parental leave legislation. In particular, wages are taxed at the household level and not at the individual level. Therefore the marginal rate of taxation depends on the household composition.

on wages in their datasets. Fourteen years after the APE reform, we are able to fill in this gap by matching two longitudinal sources (the DADS and EDP samples) containing information on career and familial background respectively. We find that the parental leave duration has a causal negative impact on later wages.

The next part presents the reform which created exogenous variations in time spent out of the labor market after the birth of a child. The econometric framework is detailed in section 3.3. Section 3.4 presents the two datasets we use in this study, as well as descriptive statistics on the matched sample. Results of the estimations are shown in section 3.5 and the last section concludes.

3.2 The 1994 parental leave reform

We take advantage of a reform that took place in France in 1994. The so called Parental Education Benefit (*Allocation Parentale d'Education*, APE thereafter) is a monthly benefit for parents who choose to temporarily reduce their labor supply after the birth of a child (beneficiaries can either work part-time or totally stop working)[2]. Parents are eligible if they have worked at least two years in the five years previous to the birth. When eligible, they receive the benefit until they come back to their previous level of labor supply. The length of the leave is up to the beneficiary, and can vary between six months and three years. Although both mothers and fathers may in theory benefit from the APE, mothers represent more than 98% of all beneficiaries (Piketty 2005). Therefore we excluded males from our study and focused exclusively on women's careers. In particular, this implies that explaining possible gender differences in wages is beyond the scope of this study.

Being an APE beneficiary does not grant any easier access to the job market once the leave is over. However, another device facilitates the return to work of young parents. The Parental Leave law (*congé parental*) applies to those who have at least one year of firm seniority at the time of the birth. Under this regime,

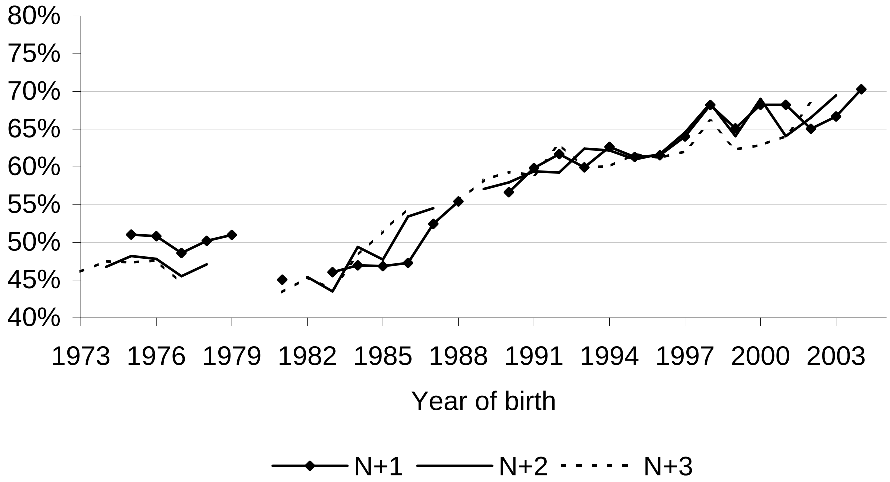
[2]This policy comes on top of mandatory maternity leave for mothers. After having given birth, a woman working under the France law is supposed to stop working during a minimum number of weeks: 10 weeks for a first or second-born, 18 weeks for a third-born child. These periods are extended in case of multiple births.

employees can ask for an unpaid leave for at most three years after the birth of a child, and the employer has to provide at the end of the leave a job similar to the one the beneficiary had before the leave. To compensate for inflation, the beneficiary's new wage must at least be equal to his old wage plus the mean increase in wages observed in the firm during the leave. In a nutshell, the APE gives financial support to working parents who want to leave their job to take care themselves of their young children, and the Parental Leave law guarantees that they won't have any trouble finding a job at the end of the leave. This Parental Leave law is in theory not related to the APE: parents can be eligible to the APE without being eligible to the Parental Leave law, and vice versa. However, in most cases parents are eligible to either both policies or none of them. As we focus on withdrawals from the labor market, we selected women who worked during the calendar year of the birth or the calendar year before. Therefore, women in our sample fullfilling APE eligibility conditions almost automatically satisfy Parental Leave eligibility conditions as well. As a consequence, slightly abusing notations, APE will refer to the combination of these two policies in the remaining of this paper.

The APE was created in 1985, and was at first available only for parents of a third-born child. Then this policy was extended to parents of a second-born child in July 1994. Judging from the number of beneficiaries, this extension has been a success from the very beginning (240 000 families with two children in 1996) until now (315 000 beneficiaries in 2002). Although the APE theoretically increases the incentives workers face to take a parental leave, these previous figures by themselves don't imply that the reform indeed affected beneficiaries' participation decision. It might be possible that they would have reduced their labor supply even in the absence of any financial compensation. In such a case, the reform would not have changed the distribution of parental leave durations. With Angrist et al.'s (1996) terminology on program evaluation, all beneficiaries would then be *always takers*, and there would be no *complier*. As we wish to use this reform as an exogenous source of variation in leave durations, we need a major part of beneficiaries to be *compliers*.

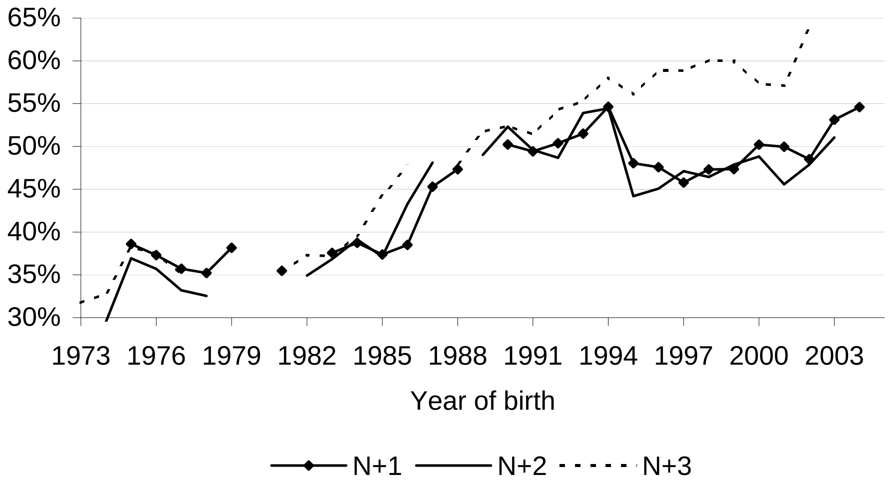
Figure 3.1, 3.2 and 3.3 illustrate that it is indeed the case and that the reform

Figure 3.1: Participation rate after the birth of the first child



Scope: Women who had a first-born child, and were present in the DADS during the year of the birth or the during previous year. N corresponds to the calendar year of the birth, N+1 is the calendar year after the birth year N. Lecture: Among women who gave birth to a first-born child in 1986 and who were working either in 1984 or in 1985, 47% were working in 1987 and 53% were working in 1988.

Figure 3.2: Participation rate after the birth of the second child



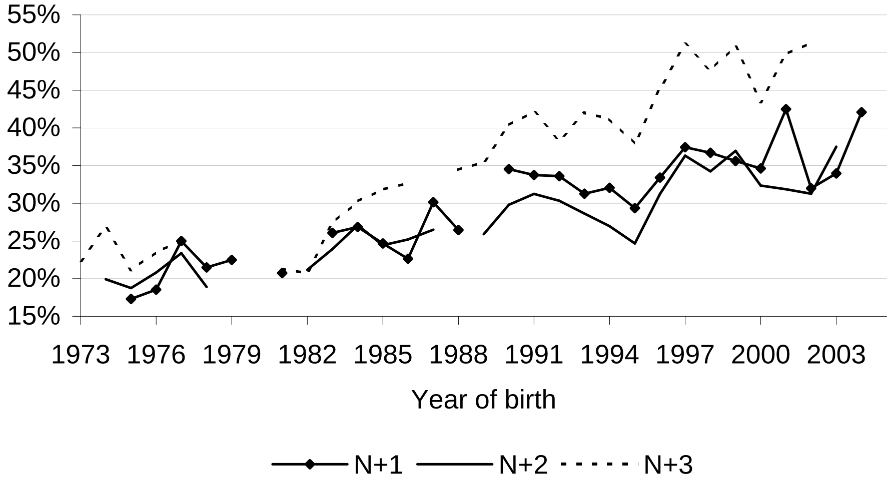
Scope: Women who had a second-born child, and were present in the DADS during the year of the birth or during the previous year. N corresponds to the calendar year of the birth, N+1 is the calendar year after the birth year N.

had a huge impact on mothers' participation rate in the three years after the birth of their second child. Participation rates 1, 2 and 3 calendar years after the birth of a first-born child are represented in Figure 3.1. Figures 3.2 and 3.3 plot the same curves for second and third-born children. A strong temporal trend is systematically visible: mothers tend to come back to work more and more frequently in the three years following the birth. The striking fact is that participation rates suddenly dropped by 10 percentage points for second-born children born after 1994, and this happened only in the first two calendar years following the birth. There is no similar decrease during the year of the birth nor 3 years after the birth. Moreover there is no such pattern with first and third-born children. This is exactly the kind of effect the APE reform was expected to produce: this reform was effective from 1994 onwards, it affected only mothers of a second-born child, and it gave incentives to withdraw during at most 3 years from the labor market. The absence of drop in the third calendar year after the birth is compatible with the APE legislation, since a withdrawal of three years after the birth implies that the return to work occurs during the third calendar year after the birth[3]. As a consequence, it is very likely that this reform accounts for most of this drop. Two points are worth emphasizing: the drop is particularly spectacular given the rising trend between 1976 and 2005, and women were fast to adapt their behavior to the new law.

Two previous papers studied the impact of the APE reform on leave duration using different datasets from ours: Piketty (2005) used Labor Force Surveys, Pailhé and Solaz (2006) worked on Family and Employers Survey. Both of them also found that the APE reform induced a significant share of eligible mothers to withdraw from the labor market in the three years following the birth of a second-born child. All this contributes to make us feel confident that our identification strategy does not rely on a weak instrumental variable.

[3]A similar pattern corresponding to the creation of APE in 1985 for parents of a third born child is present in Figure 3.3, even the drop is less clear.

Figure 3.3: Participation rate after the birth of the third child



Scope: Women who had a third-born child, and were present in the DADS during the year of the birth or during the previous year. N corresponds to the calendar year of the birth, N+1 is the calendar year after the birth year N.

3.3 The model

3.3.1 Econometric issues

When estimating the impact of parental leave on later wages with a longitudinal dataset, three problems can arise. The first issue is that unobserved time-constant individual characteristics affecting wages are likely to be correlated with some explanatory variables (for instance, ability with education). A classical way to deal with heterogeneity is to add an idiosyncratic term c_i in the wage equation. Conditioning on that unobserved effect allows the explanatory variables to be correlated with the constant component of the unobserved heterogeneity. This solves the problem of omitted variables that are constant over time. For estimation purposes, the usual procedure consists in time-demeaning the equation of interest, and then pooling all the observations for an OLS estimation on the resulting equation. This *within* estimator is consistent on a balanced panel, as long as all explanatory variables are strictly exogenous (i.e. not correlated with past, present and future values of the error term) conditional on c_i.

However careers are often discontinuous, with periods spent out of the labor market. Apart from parental leave, these can be unemployment spells, or working spells out of the private sector[4]. Working on an unbalanced panel is problematic if the selection process is non random, because not correcting for that selection may result in inconsistent estimates. Our dataset contains women who chose to work in the private sector. If the decisions to participate (year after year) in the labor market and to work in the private sector are correlated with unobserved factors affecting wages, estimations are likely to be biased. Three panel estimators have been recently suggested to take into account unobserved heterogeneity and sample selection. They all allow individual effects to be correlated with explanatory variables in both the selection and primary equations[5]. Kyriazidou's (1997) estimator relies on individuals who have "close" selection effects in two different time periods. Differencing these two observations removes at the same time the individual and selection effects. Therefore the selection effect remains an unknown function, and requires no assumption. On the other hand, Rochina-Barrachina (1999) and Wooldridge (1995) parametrize this selection bias. The former removes the unobserved effect by differencing observations for individuals whose wage is observed twice. The latter applies the transformation proposed by Mundlak (1978) to deal with unobserved heterogeneity, and follows Heckman (1976) to correct for selection bias. He then estimates the wage equation in levels.

Apart from heterogeneity and sample selection, the third issue arising in our study is that parental leave length may suffer from measurement error and endogeneity. As the DADS covers only the private sector, the length l_{it} measured in the DADS may overestimate the actual length of withdrawal from the labor market if the mother's first job after the birth is not in the private sector. Moreover, endogeneity may stem from the link between l_{it} and motivation through the trade off

[4]The DADS covers workers in the private sector, and doesn't contain civil servants and independent workers. See section 3.4.

[5]Dustmann and Rochina-Barrachina (2007) survey these estimators in detail. They compare the exact set of assumptions needed for each estimator, and provide a discussion on their respective pros and cons/advantages and drawbacks. In particular they point out that Kyriazidou (1997) requires a "*conditional exchangeability* assumption which may be rather restrictive in practical applications".

made between time spent at the workplace and at home after the birth. This kind of correlation is indeed allowed in the fixed effect specification, as long as motivation keeps constant over time. But this assumption might be too strong, as some mechanisms could induce variations in motivation during the career. Changes in personal life like getting married or having children may increase the will to have spare time devoted to family, and thus lower motivation. Hence l_{it} might be linked to contemporaneous idiosyncratic changes in wages. Besides, even if motivation were constant over time, the three aforementioned estimators require strict exogeneity for the length l_{it}, which is unlikely in our context. For instance, negative exogenous shocks to wages in the past may be related to poor work conditions, and thus affect the choice of parental leave length today. Dustmann and Rochina-Barrachina (2007) show that the previous estimators can be adapted to cases where strict exogeneity fails.

Our econometric specification is derived from Semykina and Wooldridge (2010). Their procedure is based on Wooldridge's (1995) estimator, and further allows some explanatory variables to be endogenous. Therefore it takes into account unobserved heterogeneity, endogenous variables, and corrects for selection bias while working on an unbalanced panel. Jäckle (2007) implements this method when studying the impact of health status on wages, while Pouget et al. (2007) focus on unemployment duration.

3.3.2 No selection effect

Let's first ignore sample selection issues, and suppose that we work on a balanced panel. The equation of interest is the following:

$$y_{it} = x_{it}\alpha + l_{it}\beta + c_i + u_{it}, \quad t = 1, \dots, T \tag{3.1}$$

y_{it} is the (log of the) annual wage of individual i in year t divided by the number of days worked during that year. x_{it} are time-varying individual characteristics affecting the wage (age, etc.), which are supposed to be strictly exogenous conditional on c_i. l_{it} is the length (in years) of the withdrawal after the birth of the second-born child: it is equal to 0 before the second birth, and to the actual length after the birth. c_i represents time-constant factors like ability or motivation. u_{it} is the error term, summing up all time-varying unobserved variables which determine wages. c_i can be arbitrarily correlated with x_{it} and l_{it}. For estimation purposes, one can implement a fixed-effect transformation (FE) to remove c_i, and then run an OLS estimation on the time-demeaned equation. This procedure gives consistent estimates on a balanced panel if x_{it} and l_{it} are strictly exogenous (i.e. not correlated with $u_{it'}$ for all t') conditional on c_i.

As explained above, it is likely that strict exogeneity of l_{it} will fail. A possible remedy is to find instrumental variables z_{it} sufficiently correlated with l_{it} and strictly exogenous conditional on c_i (z_{it} contains all strictly exogenous variables x_{it}, plus other strictly exogenous variable(s) correlated with l_{it}). The procedure consists in time-demeaning equation (3.1) like in FE estimations and then applying a two stage least squares estimation (2SLS) to the time-demeaned equation. This FE-2SLS method produces consistent estimates on a balanced panel.

Semykina and Wooldridge (2010) show that it can be applied to unbalanced panels, under a more restrictive condition. Let s_{it} be a binary variable equal to 1 if (y_{it}, x_{it}, z_{it}) is observed and 0 otherwise. Then in addition to the usual rank conditions, the following assumption is needed:

$$\mathbb{E}(u_{it}|z_i, s_i, c_i) = 0, \quad t = 1, \ldots, T \tag{3.2}$$

where $z_i = (z_{i1}, ..., z_{iT})$ and $s_i = (s_{i1}, ..., s_{iT})$. A major feature of FE-2SLS is that no restriction is imposed on the relationship between s_i and (c_i, z_i). In particular, it allows attrition to be correlated with unobserved heterogeneity, which will be the case if some constant characteristics determining wages also have an impact on selection. Given the strict exogeneity of (z_i, c_i), (3.2) automatically holds in the two polar cases where selection is totally random (i.e. not correlated with observed and unobserved determinants of wages) or completely determined by (z_i, c_i). However neither of these situations seems likely to occur. For instance, one's level of education is a plausible candidate to explain both participation and wages, which rules out randomness. Moreover, assuming that all possible parameters related to the decision to participate into the labor market are included in (z_i, c_i) seems unrealistic: among other things, it would imply that there would be no unobserved time-varying variable influencing participation. But even if none of these two extreme situations holds, (3.2) remains valid as long as determinants of s_{it} not included in (z_i, c_i) are not part of the unexplained changes in wages $u_{it'}$ for all t'. This last assumption seems too strong in our context, since idiosyncratic shocks on wages in year t' could affect the decision to participate during year t ($t > t'$). Annex 3.8.3 details a procedure to test whether condition (3.2) holds.

If condition (3.2) indeed fails, FE-2SLS cannot be used to estimate (3.1). In such a case, Semykina and Wooldridge (2010) provide an estimation strategy which overcomes the limitation faced by the above FE-2SLS on unbalanced panels. Their method allows for endogeneity of some explanatory variables, and constant unobserved heterogeneity to be correlated with explanatory and instrumental variables in both the selection and wage equations. Selection $s_{it'}$ can be correlated with u_{it} (for all t and t'), and contemporaneous selection bias is corrected for. This procedure requires the instrumental variables z_{it} to be observed only when $s_{it} = 1$, and to be strictly exogenous conditional on c_i. However z_{it} and c_i can be arbitrarily correlated, so the most attractive feature of FE is also present here. Eventually additional parametric assumptions are needed. The remaining of this section presents the modeling of the selection and wage equations.

3.3.3 The selection process

The selection process determining participation into the labor market during year t is specified using a probit model:

$$\begin{cases} s_{it} = \mathbb{1}[Z_{it}\gamma + d_i + v_{it} > 0] \\ v_{it}|Z_i, d_i \sim \mathcal{N}(0,1), \quad t = 1, \ldots, T \end{cases} \tag{3.3}$$

The selection indicator s_{it} equals 1 if individual i worked at least one day during year t. $Z_i = (Z_{i1}, ..., Z_{iT})$, where Z_{it} is a vector containing x_{it} and at least one other variable not present in equation (3.1). The econometric model is theoretically identified without any exclusion variable, but in that case identification relies solely on the non linearity of equation (3.3). The specification becomes more convincing when there is at least one variable affecting selection and not wages. The fixed effect d_i sums up all persistent heterogeneity which could explain the propensity of individual i to select in or out of the sample. If we ignore d_i and consider it part of the error term, then the errors terms are automatically serially correlated. According to Semykina and Wooldridge (2010), estimation then imposes further assumptions which are far too unrealistic. We suppose that d_i and Z_i can be correlated, as it is likely to be the case. FE cannot be applied here to take d_i into account, since equation (3.3) is not linear. Mundlak (1978) proposed a way to deal with d_i without time-demeaning. He modeled the relationship between d_i and the explanatory variables Z_i as follows:

$$\begin{cases} d_i = \mu + \overline{Z_i}\delta + a_i \\ a_i|Z_i \sim \mathcal{N}(0, \tau^2) \end{cases} \tag{3.4}$$

This equation writes d_i as the sum of a term correlated with Z_i and a part which by construction is independant of Z_i. In all generality, no restriction should be imposed on the linear projection of d_i on Z_i, and therefore the time-averaged variables $\overline{Z_i} = (Z_{i1} + ... + Z_{iT})/T$ in equation (3.4) should be replaced by Z_i (this more flexible specification was suggested by Chamberlain 1980). However, d_i does not vary over time, and it seems legitimate to restrict the projection of d_i on Z_i to time-constant functions of Z_i. Mundlak's (1978) specification amounts to choosing a simple time-invariant function by imposing the same coefficient for $Z_{it'}$ at all periods t'. In other words, (3.4) assumes that all interactions between d_i and Z_i are captured by the time average $\overline{Z_i} = (Z_{i1} + ... + Z_{iT})/T$. Both approaches are valid within our framework, we chose to present Mundlak's here because it conserves on degrees of freedom. Unlike in the FE transformation, Z_i can contain time-constant variables like education.

Plugging (3.4) into equation (3.3) leads to:

$$\begin{cases} s_{it} = \mathbb{1}[\mu + Z_{it}\gamma + \overline{Z_i}\delta + w_{it} > 0] \\ w_{it}|Z_i \sim \mathcal{N}(0, 1+\tau^2), \quad t = 1, \dots, T \end{cases} \tag{3.5}$$

where $w_{it} = a_i + v_{it}$. In fact, the effect of d_i in equation (3.3) or the variance of w_{it} can be allowed to vary over time. Therefore a more general specification of the selection process is (after a rescaling of the error term):

$$\begin{cases} s_{it} = \mathbb{1}[\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t + w_{it} > 0] \\ w_{it}|Z_i \sim \mathcal{N}(0, 1), \quad t = 1, \dots, T \end{cases} \tag{3.6}$$

3.3.4 The wage equation

First let's recall that the fixed effect transformation was applied to remove the unobserved heterogeneity c_i from equation (3.1) before running a 2SLS estimation. The residual resulting from the time-demeaning was a function of $u_{it'}$ for all t'. Doing so, correlation between the selection indicator in period t and $u_{it'}$ (for all t') became a problem, whereas only contemporaneous selection mattered in equation (3.1). Once again, Mundlak's (1978) device can be used to avoid time-demeaning. The relationship between c_i and the strictly exogenous variables z_i is supposed to take the following form:

$$\begin{cases} c_i = \eta + \overline{z_i}\theta + b_i \\ \mathbb{E}(b_i|z_i) = 0 \end{cases} \tag{3.7}$$

This specification assumes that c_i depends on z_i only through the time average $\overline{z_i} = (z_{i1} + ... + z_{iT})/T$. Note the no assumption is made on the law of $b_i|z_i$. The wage equation (3.1) can be rewritten using (3.7):

$$y_{it} = x_{it}\alpha + l_{it}\beta + \eta + \overline{z_i}\theta + b_i + u_{it}, \quad t = 1, \dots, T \tag{3.8}$$

To highlight the correction for contemporaneous selection bias, we can write:

$$\begin{cases} y_{it} = x_{it}\alpha + l_{it}\beta + \eta + \overline{z_i}\theta + \mathbb{E}(b_i + u_{it}|z_i, s_{it}) + e_{it} \\ \mathbb{E}(e_{it}|z_i, s_{it}) = 0, \quad t = 1, \dots, T \end{cases} \tag{3.9}$$

One important feature of (3.9) is that it is silent on correlation between e_{it} and $s_{it'}$ for $t \neq t'$. Therefore we don't have to take into account selection in other periods, even if this selection indicator is correlated with e_{it}. In other words, selection does not have to be strictly exogenous. If we knew $\mathbb{E}(b_i + u_{it}|z_i, s_{it})$, applying pooled 2SLS to (3.9) would give consistent estimates of the parameters. In fact we only need to compute $\mathbb{E}(b_i + u_{it}|z_i, s_{it} = 1)$, because we do not observe (y_{it}, x_{it}) when $s_{it} = 0$. The following linearity assumptions

$$\begin{cases} \mathbb{E}(u_{it}|z_i, w_{it}) = \mathbb{E}(u_{it}|w_{it}) = \rho_t w_{it}, \quad t = 1, \dots, T \\ \mathbb{E}(b_i|z_i, w_{it}) = \mathbb{E}(b_i|w_{it}) = \psi_t w_{it}, \quad t = 1, \dots, T \end{cases} \tag{3.10}$$

are classical and imply that the functions of w_{it} which best fit $\mathbb{E}(u_{it}|w_{it})$ and $\mathbb{E}(b_i|w_{it})$ are linear. (3.10) automatically holds in the special case where (u_{it}, w_{it}) (resp. (b_i, w_{it})) follow a bivariate normal distribution. The slopes of the linear fits are allowed to differ from one time period to another. Noting $\Psi_t \equiv \rho_t + \psi_t$, we use (3.10) and the law of iterated expectations to get:

$$\begin{aligned} \mathbb{E}(b_i + u_{it}|z_i, s_{it} = 1) &= \mathbb{E}(b_i + u_{it}|z_i, w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t) \\ &= \mathbb{E}(b_i + u_{it}|w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t) \\ &= \frac{\mathbb{E}\Big[(b_i + u_{it}) * \mathbb{1}(w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t)\Big]}{\mathbb{P}\big[w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t\big]} \\ &= \frac{\mathbb{E}\Big[\mathbb{E}(b_i + u_{it}|w_{it}) * \mathbb{1}(w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t)\Big]}{\mathbb{P}\big[w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t\big]} \\ &= \Psi_t \frac{\mathbb{E}\Big[w_{it} * \mathbb{1}(w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t)\Big]}{\mathbb{P}\big[w_{it} > -\mu_t - Z_{it}\gamma_t - \overline{Z_i}\delta_t\big]} \\ &= \Psi_t \frac{\phi(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t)}{\Phi(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t)} \end{aligned} \tag{3.11}$$

where ϕ and Φ are respectively the probability density and cumulative distribution functions of a standard normal law. It shows that under (3.10), $\mathbb{E}(b_i + u_{it}|z_i, s_{it} = 1)$ is proportional to the inverse Mills ratio

$$\lambda_{it}(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t) = \frac{\phi(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t)}{\Phi(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t)}$$

Running the probit regression (3.6) (for each period t separately) provides an estimate of this ratio $\widehat{\lambda_{it}}(\mu_t + Z_{it}\gamma_t + \overline{Z_i}\delta_t) = \lambda_{it}(\widehat{\mu_t} + Z_{it}\widehat{\gamma_t} + \overline{Z_i}\widehat{\delta_t})$.

Eventually, the estimation strategy is the following:

- Compute the estimate of the inverse Mills ratio $\widehat{\lambda_{it}}$ for period t from equation (3.6), $\quad t = 1, \ldots, T$
- Replace $\mathbb{E}(b_i + u_{it}|z_i, s_{it})$ by $\Psi_t\widehat{\lambda_{it}}$ in equation (3.9), and estimate (3.9) on the selected sample ($s_{it} = 1$) by pooled 2SLS. The instrumental variables are 1, z_{it}, $\overline{z_i}$ and $\widehat{\lambda_{it}}$. The estimators' variance-covariance matrix needs to be computed according to the formula given in 3.8.2.

The presence of at least one exclusion variable in the probit estimations guarantees that even if the inverse Mills ratio is well approximated by a linear function on a large part of its range, there won't be any collinearity issue in the second step.

In cases where condition (3.2) fails, this procedure corrects for selection bias and endogeneity of l_{it}. Moreover, it allows unobserved heterogeneity to be correlated with explanatory variables, both in the selection and primary equations. It also allows for correlation between the idiosyncratic errors in the two equations. Both error terms can be serially correlated and heteroscedastic. Joint normality of the error terms is not required: the error term in the selection equation is supposed to be normally distributed, while there is only a linearity assumption on the conditional mean of u_{it}.

Eventually, using Mundlak's (1978) device instead of the usual FE transformation in the equation of interest allows us to add time-constant variables like education in equation (3.1). Obviously, their effect on wages (given by the vector of parameters α) is not identified in this procedure: it cannot be distinguished from the effect of unobserved time-constant variables passing through the time-averaged $\overline{z_i}$ after the Mundlak's (1978) transformation. However, adding these variables is likely to capture a greater part of the unobserved heterogeneity constant over time in (3.7). Therefore, even if these coefficients cannot be causally interpreted, it is still an improvement compared to a situation with no time-constant variable in the wage equation.

3.4 Data

We use information from two sources. The *Déclarations Annuelles de Données Sociales* (DADS thereafter) is a large-scale administrative dataset containing information on each employee subject to French payroll taxes. It basically includes all employees or self-employed persons working in private and state-owned firms. Only civil servants and independent workers are not present in the DADS. The DADS gathers yearly reports filled by employers. An observation consists in a unique individual-firm-year triplet. Each observation contains the number of days worked by the individual in the establishment during the calendar year, as well as the first day of the first spell of employment and the last day of the last spell of employment during that calendar year. It also provides us with date of birth, sex, occupation, full-time/part-time status[6], the total net nominal earning and the annualized gross nominal earnings for the individual in that year. We exploit an extract of the DADS, covering all women born in October of even-numbered years. We follow them between 1976 and 2005 (except for 1981, 1983 and 1990, because the extraction of the DADS was not made in these three years).

The Permanent Demographic Sample (*Echantillon Démographique Permanent*, or EDP) provides us with general information on individuals. This longitudinal dataset covers all French women and men born on one of the first four days of October. It compiles 1968, 1975, 1982, 1990 and 1999 census data with information from register of births, marriages and deaths from 1968 to 2005. In particular, it contains for each individual in the sample the dates of their children's birth.

The EDP and DADS use the same individual identifier NIR (a 13 digit number) which allows us to match these two datasets. However, we first had to exclude persons not born in France, because their identifier was not built with the same

[6]A dummy variable for part- or full-time employment is present between 1976 and 2005. However we do not know what fraction of a full-time job a part-time employee works. The annual number of paid hours is available only from 1994 on, so we can't use this variable in our model. All we can do is provide descriptive statistics for the post-reform period. This lack of information in our dataset led us to focus exclusively on total withdrawals from the labor market, although the APE also allows for part-time jobs.

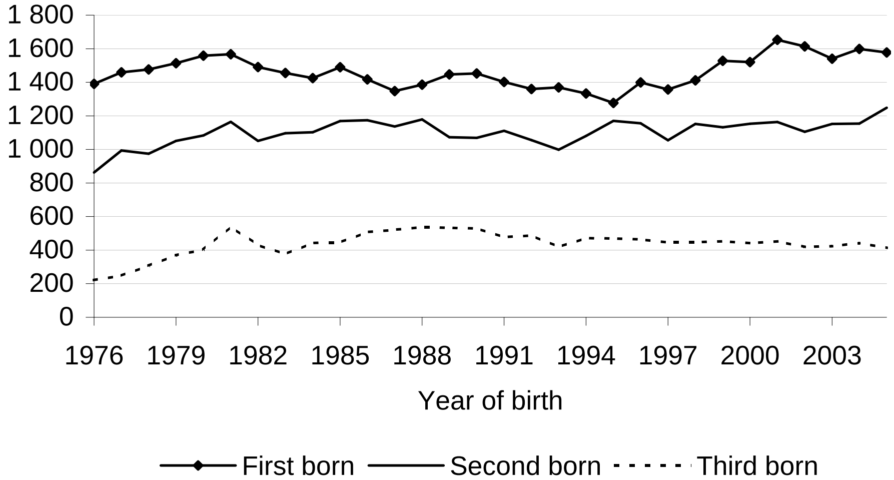
algorithm in the two sources. This removed 15% of individuals in the EDP and 10% of observations in the DADS. We also excluded DADS observations with an obviously wrong NIR (containing letters instead of numbers). When we matched these two samples, we selected women born on one of the first four days of October of even-numbered years. These women have worked at least one day in their life in the private sector. The matched sample contains 99,505 women and 1,285,407 observations. By construction, the EPD is representative of the French population, while the DADS may not be. Annex 3.8.1 provides information on possible sample selection issues by checking whether women's observable characteristics in the matched sample systematically differ from those of EDP.

Figure 3.4 represents the number of births by year. It shows that the reform was not followed by an increase in the number of second-born children. Therefore the decision to have a second child was not severely affected by the APE reform. This is crucial in our identification strategy, which imposes that women who gave birth to a second child before and after the reform had similar characteristics (so that the sample selection bias created by selecting mothers with at least two children is constant over time). However, the total number of birth increases in the 1990ies, mainly driven by the rise in first-born children. Piketty (2005) noted the same phenomena, and argued that most of this increase was probably due to better macroeconomic conditions. He estimated that the APE reform explained only a small part of the increase between 1994 and 2001. See Piketty (2005) for more information on that specific subject.

Figure 3.5 is devoted to participation rate in the labor market at different ages. Each curve refers to a given birth cohort (1950, 1960, 1970 and 1980), and plots the proportion of women who appeared in the DADS between 1976 and 2005. The progressive rise of female labor market participation rate observed in most developed countries certainly accounts for the increase observed between the 1950 and 1980 cohorts.

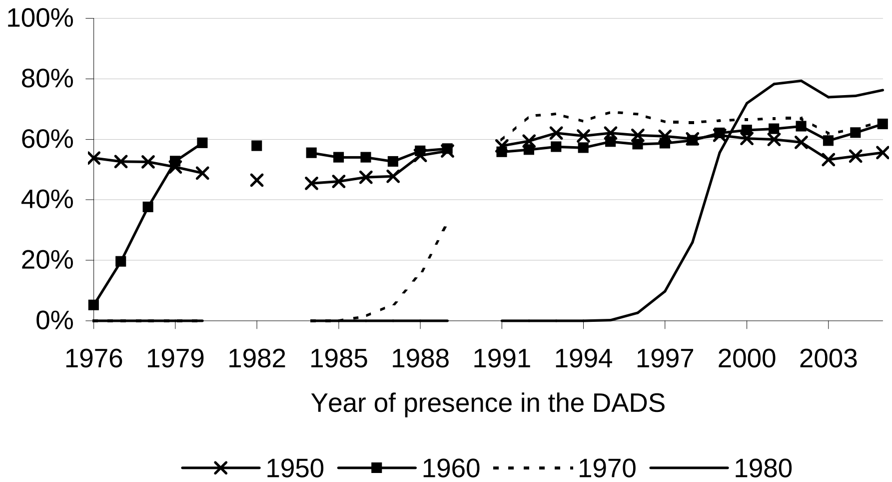
After having given birth, a woman working under the French law is supposed to stop working during at least a given number of weeks, corresponding to the mandatory maternity leave. In most cases the actual length of the withdrawal

Figure 3.4: Number of births per year



Notes: Number of births observed each year, with the mother belonging to the EDP/DADS sample.

Figure 3.5: Participation rate by birth cohort



Scope: Women present in the EDP/DADS sample. Lecture: 80% of the 1980 cohort were present in the DADS (i.e. worked a least one day in the private sector) in 2002.

is greater than this minimum leave. This length is a key variable in our study, as we wish to estimate the marginal impact of the withdrawal on later wages. Participation rates 1, 2 and 3 calendar years after the birth of a first-born child are represented in Figure 3.1. Figures 3.2 and 3.3 plot the same curves for second and third-born children. A strong temporal trend is systematically visible: mothers tend to come back to work more and more frequently in the three years following the birth. The striking fact is that participation rates suddenly dropped by 10 percentage points for second-born children born after 1994, and this happened only in the two calendar years following the birth. There is no similar decrease during the year of the birth nor 3 years after the birth. Moreover there is no such pattern with first and third-born children. As the APE reform in 1994 affected only mothers of a second-born child, and gave incentives to withdraw during at most 3 years from the labor market, it is very likely that this reform accounts for most of these drops[7]. Two points are worth emphasizing: the drops are particularly spectacular given the rising trend between 1976 and 2005, and women were fast to adapt their behavior to the new law. Piketty (2005) found similar results using French Labor Force Surveys.

Figures 3.6, 3.7, and 3.8 represent the cumulative frequency of the number of years spent out of the labor market after the birth of a child. Separate curves are plotted depending on whether the child was born before or after the APE reform took place in 1994 (note that 1994 is only a milestone in Figures 3.6 and 3.8, because the APE reform in 1994 did not change incentives for mothers of a first- or third-born child). Mothers tend to come back to work more often and more rapidly when their child is born after 1994. This result is in line with the rising trend in participation rates after the birth observed in Figures 3.1 to 3.3. It certainly stems from the general change in women's (and here especially mothers') behavior toward the labor market, often symbolized by the rise in female participation rate (Blanchet and Pennec 1997). The gap between the pre-reform and post-reform curves is

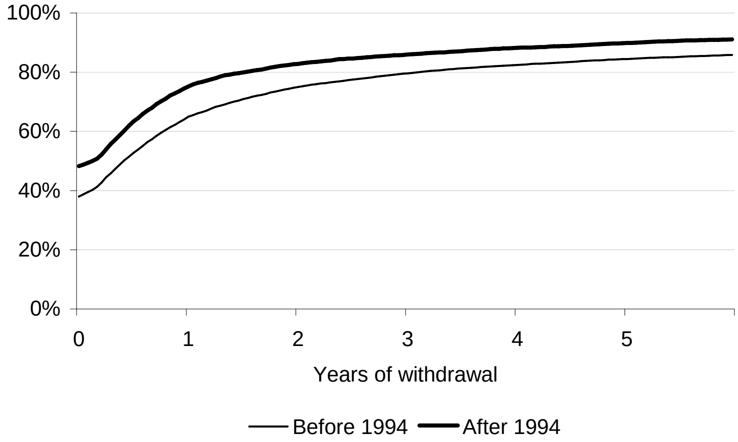
[7] The absence of drop in the third calendar year after the birth is not problematic, since a withdrawal of three years after the birth implies that the return to work occurs during the third calendar year after the birth.

roughly constant after the birth of a first-born child. Therefore the behavior change seems to have evenly affected all working mothers when it comes to parental leave duration. A similar conclusion can be drawn from Figure 3.8 for mothers of three children. Moreover one can notice a negative shift for both curves between zero and three years of withdrawal. Such a negative shift is the kind of effect that the APE could create, since the APE gives incentives to delay the return to work during the first three years. As the APE was available for mothers of a third-born child since 1985, it could indeed have affected the two curves and thus is a plausible candidate to explain (at least part of) this downward change. Figure 3.7 shows a similar shift for mothers of a second-born child, but only for the post-reform curve. Furthermore the gap is not strongly marked for short withdrawals (less than 6 months), and then becomes wider until the spell reaches three years. All this strengthens the hypothesis that the APE caused these shifts, because mothers of a second-born child became eligible to the APE in 1994, and withdrawals under the APE legislation can vary between 6 months and three years. Once again, these observations are in line with previous studies (Pailhé and Solaz 2006). It is worth noticing that these shifts occur around three years of withdrawal. Since the APE is available until the third anniversary of the child, it might imply that a significant proportion of APE beneficiaries choose to return to work right before the maximum length of three years elapses.

Giving birth may affect subsequent career path in different ways. There may be a wage penalty associated with the simple fact of having a baby[8]. We would then observe a decrease in mothers' wages after the birth. In all generality, this decrease could be time-constant, or could vary with the number of years since the birth. On top of this "scar" effect, the duration of the withdrawal from the labor market after the birth may also impact wages. Figure 3.9 pictures the mean wage between 1976 and 2005 of women who gave birth in 1993 (either of a first-, second- or third-born child). Women who left the labor market only during the mandatory maternity leave have a higher wage after the birth than mother who withdrew longer. The gap appears in 1995 and is roughly constant afterwards, whereas there

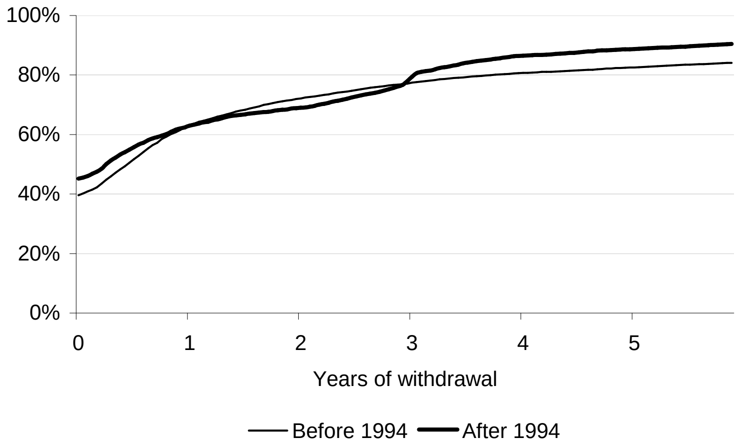
[8] See Felfe (2006).

Figure 3.6: Cumulative spell duration after the birth of the first child



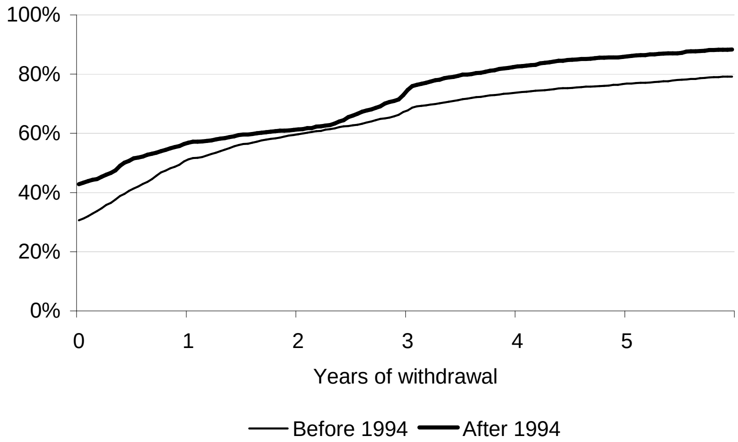
Scope: Women who had a first-born child between 1976 and 1999 included, and were present in the DADS during the year of the birth or during the previous year. The curves represent the cumulative frequency of the length of spell out of the labor market after the birth. The plain line is for women who gave birth before July 1994, the bold line for woman who gave birth between July 1994 and December 1999. As we have information until 2005, the length of withdrawal is defined up to 6 years in the latter curve.

Figure 3.7: Cumulative spell duration after the birth of the second child



Scope: Women who had a second-born child between 1976 and 1999 included, and were present in the DADS during the year of the birth or during the previous year. The curves represent the cumulative frequency of the length of spell out of the labor market after the birth. The plain line is for women who gave birth before July 1994, the bold line for woman who gave birth between July 1994 and December 1999. As we have information until 2005, the length of withdrawal is defined up to 6 years in the latter curve.

Figure 3.8: Cumulative spell duration after the birth of the third child

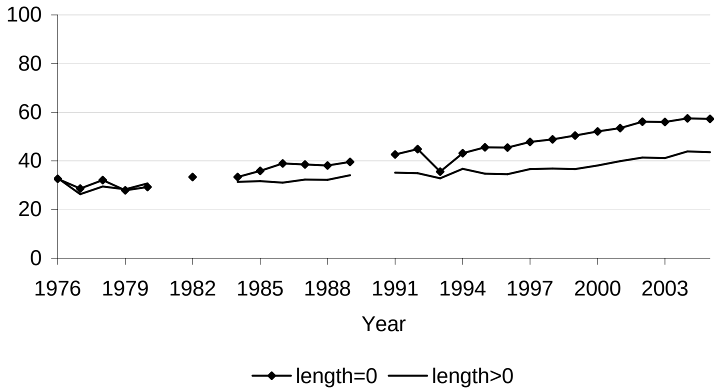


Scope: Women who had a third-born child between 1976 and 1999 included, and were present in the DADS during the year of the birth or during the previous year. The curves represent the cumulative frequency of the length of spell out of the labor market after the birth. The plain line is for women who gave birth before July 1994, the bold line for woman who gave birth between July 1994 and December 1999. As we have information until 2005, the length of withdrawal is defined up to 6 years in the latter curve.

is no significant difference in wages before the birth. This pattern is not specific to the 1993 birth cohort (results not presented here, and available upon request), and hence seems to be quite general. At this point, it is not possible to interpret this as a causal relationship running from the withdrawal duration to a decrease in wages. There may be other characteristics negatively affecting wages after the birth and common to all women who chose to withdraw longer.

Figure 3.10 focuses on mothers who gave birth to a second-born child. All births occurred in 1996, so these mothers were potentially eligible to APE. A gap similar to Figure 3.9 is visible after the birth, its magnitude is constant until 2005. However this gap does not appear right after the birth, but rather three years after. This may be due to a composition effect. Piketty (2005) argues that low wages women tend to withdraw longer than high wages women when using the APE. Hence we would observe relatively more high wages women working in 1997 and 1998 among women who temporarily withdraw from the labor market after the birth. This would explain why wages are not decreasing (and are even increasing) in those two years, relatively to mothers whose withdrawal did not exceed the mandatory maternity

Figure 3.9: Mean daily wage per year, for women who gave birth in 1993



Notes: Annual wage divided by the number of days worked during the year, in €2005. Scope: Women who gave birth in 1993 to a first-, second- or third-born child. "length=0" covers women who withdrew from the labor market only during the mandatory maternity leave after the birth. "length>0"corresponds to women who took a break longer than the mandatory maternity leave.

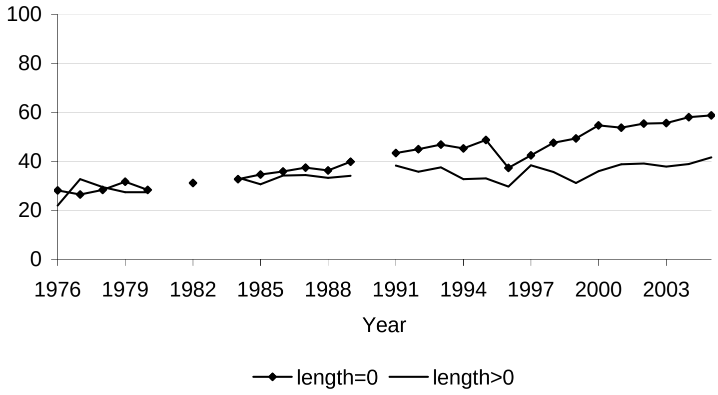
leave.

A peculiar pattern is visible in both Figures 3.9 and 3.10. There is a drop in wages of about 20% the year of the birth for women who took only the mandatory maternity leave. Wage growth rates do not seem to differ before and after the birth. This decrease is common to all birth cohorts, and does not depend on whether women gave birth to a first-, second- or third-born child. This may partly be due to a (permanent) shift from full-time to part-time employment after the birth for some of these mothers.

3.5 Results

We implement the estimation strategy described in section 3.3. The estimation sample is composed of women who gave birth to a second child between 1986 and 2002, and their wages are observed between 1984 and 2005. l_{it} is the length (in years) of the withdrawal from the labor market following the birth of the second-born child. Our instrumental variable z_{it} correcting for the endogeneity of l_{it} is whether this birth occurred before or after the APE reform in July 1994. The

Figure 3.10: Mean daily wage per year, for women who gave birth to a second-born in 1996



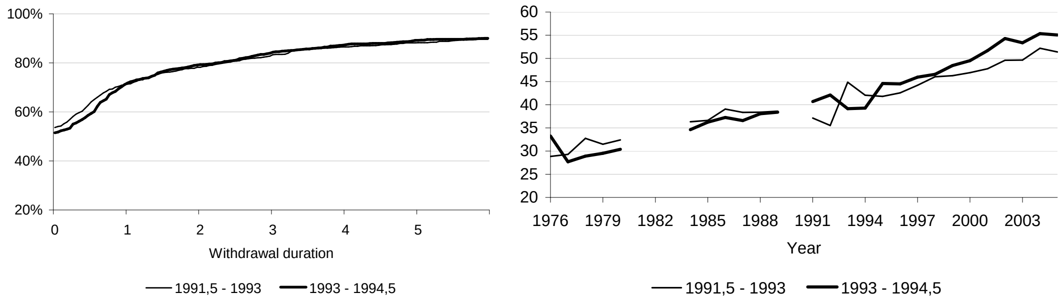
Notes: Annual wage divided by the number of days worked during the year, in €2005. Scope: Women who gave birth to a second-born child in 1996. "length=0" covers women who withdrew from the labor market only during the mandatory maternity leave after the birth. "length>0" corresponds to women who took a break longer than the mandatory maternity leave.

first exclusion variable in the selection equation is a dummy equal to one if the woman has at least a child under the age of three. Several authors used a similar exclusion variable. Local unemployment rate is also used as an exclusion variable. Region of residence and a part-time job indicator are among explanatory variables in the wage equation. As this latter variable might be considered as endogenous when determining wages, we systematically run estimations with and without this variable in the set of regressors. Finally, covariates common to both wage and selection equations are age, square age, number of children, and annual dummies as a proxy for macro economic environment[9].

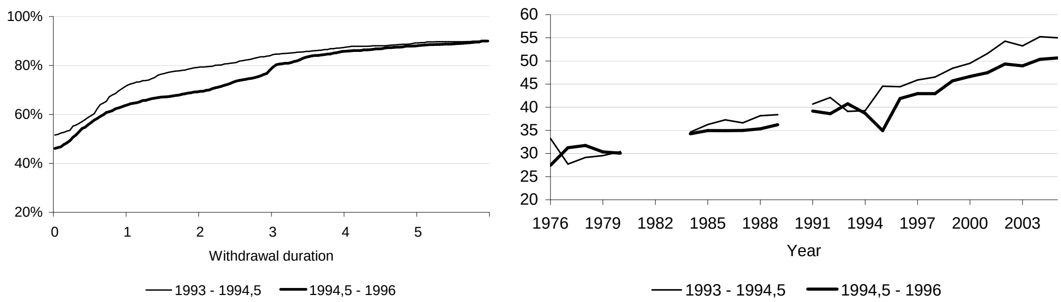
The remaining of this section presents the results of different estimations on this sample. Table 3.1 presents the FE estimation results of equation (3.1). This estimation allows for unobserved heterogeneity to be correlated with all explanatory

[9]Individual variables constant over time cannot be included in the regression, because their effect cannot be separated from the effect of the constant unobserved heterogeneity. Hence we don't include year of birth, education and socio-professional group. Sex is not included either since our sample contains only women.

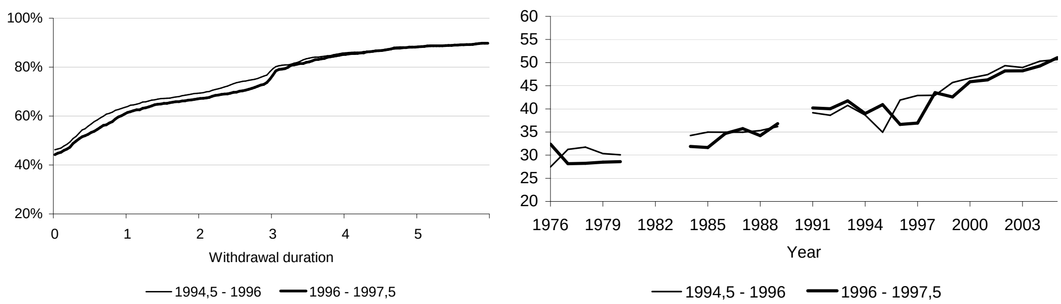
Figure 3.11: Withdrawal duration and wages, for mothers of a second-born child



(a) Child born in (July 1991, December 1992) vs. (January 1993, June 1994)



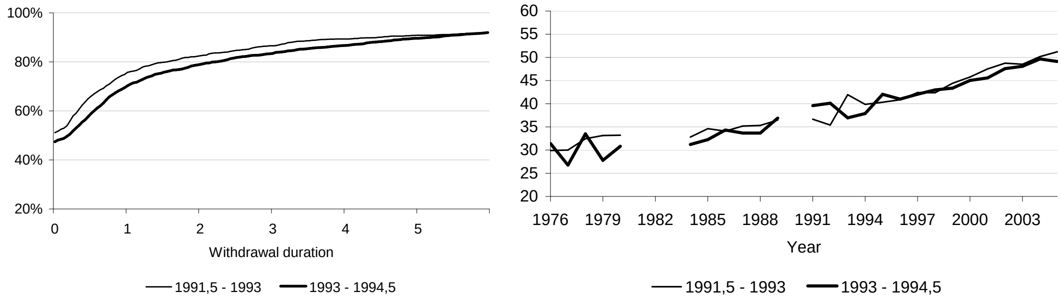
(b) Child born in (January 1993, June 1994) vs. (July 1994, December 1995)



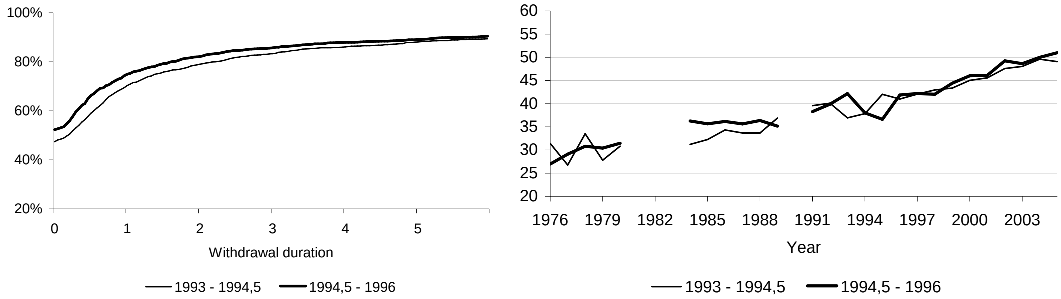
(c) Child born in (July 1994, December 1995) vs. (January 1996, June 1997)

Notes: Figure 3.11 includes women who had a second-born child between July 1991 et June 1997. The cumulative frequency of withdrawal duration is pictured on the left. Their mean daily wage in €2005 is represented on the right.

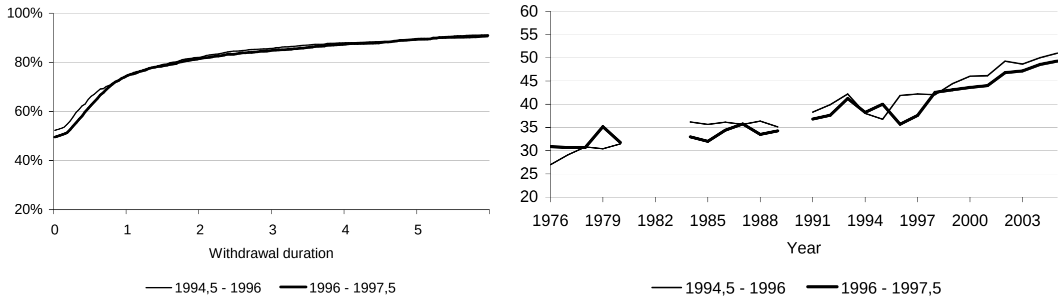
Figure 3.12: Withdrawal duration and wages, for mothers of a first-born child



(a) Child born in (July 1991, December 1992) vs. (January 1993, June 1994)



(b) Child born in (January 1993, June 1994) vs. (July 1994, December 1995)



(c) Child born in (July 1994, December 1995) vs. (January 1996, June 1997)

Notes: Figure 3.12 includes women who had a first-born child between July 1991 et June 1997. The cumulative frequency of withdrawal duration is pictured on the left. Their mean daily wage in €2005 is represented on the right.

Table 3.1: FE estimation of equation (3.1)

	Dependant variable *wage* (1)	(2)
$length$	-0.030***	-0.032***
	(0.0019)	(0.0020)
age	0.065***	0.084***
	(0.0021)	(0.0022)
$age2$	-0.055***	-0.082***
	(0.0028)	(0.0030)
$nbchild$	-0.049***	-0.11***
	(0.0034)	(0.0035)
$part-time$	-0.56***	
	(0.0036)	
$intercept$	2.40***	1.96***
	(0.053)	(0.056)
$temporal\ FE$	yes	yes
$regional\ FE$	yes	yes

Notes: Total sample consists of women in the EDP/DADS matched sample, who gave birth to a second-born child between 1986 and 2002. Their wage is observed more than one year for 17 389 of them. Wages are observed 204 177 times between 1984 and 2005. $part-time$ is a dummy variable equal to 1 if and only if the woman worked part-time during the corresponding year. The dependent variable $wage$ is the log of the average daily wage earned in the corresponding year, in €2005.

variables, and ignores potential sample selection issues. Moreover strict exogeneity is assumed for all variables. Whether or not including a part-time job indicator in the covariates, one year of withdrawal from the labor market is associated with a 3% decrease in wages.

Unobserved heterogeneity and endogeneity are taken into account in a FE-2SLS estimation. On the other hand, sample selection is not accounted for and it might lead to inconsistent estimates. Columns 1 and 2 of Table 3.2 show the results of the first step estimation. The instrumental variable z has a large (+0.10 year) and highly significant effect on withdrawal leave duration, in line with statistics presented in section 3.4. The second step of the estimation (columns 3 and 4) indicates that parental leave duration has a negative and significant impact on

Table 3.2: FE-2SLS estimation

	First stage		Second stage	
	length		*wage*	
	(1)	(2)	(3)	(4)
length			-0.15**	-0.21***
			(0.064)	(0.069)
z	0.11***	0.11***		
	(0.0084)	(0.0084)		
age	-0.0053**	-0.0045*	0.065***	0.085***
	(0.0026)	(0.0026)	(0.0021)	(0.0023)
age2	0.032***	0.032***	-0.052***	-0.078***
	(0.0036)	(0.0036)	(0.0032)	(0.0034)
nbchild	0.26***	0.26***	-0.014	-0.059***
	(0.0046)	(0.0046)	(0.019)	(0.020)
$part-time$	0.0083*		-0.56***	
	(0.0044)		(0.0036)	
intercept	0.076	-0.10	2.27***	1.93***
	(0.12)	(0.066)	(0.10)	(0.059)
$temporal\ FE$	yes	yes	yes	yes
$regional\ FE$	yes	yes	yes	yes

Notes: Total sample consists of women in the EDP/DADS matched sample, who gave birth to a second-born child between 1986 and 2002. Wages are observed 204 177 times between 1984 and 2005. The dependent variable *wage* is the log of the average daily wage earned in the corresponding year, in €2005.

wages. Its magnitude is larger than in the FE estimation: wages decrease by 15% with each year away from the labor market (column 3).

Results of the procedure testing for selection bias are presented in section 3.8.3. They show that there is indeed a significant selection bias (the null hypothesis of no contemporaneous selection bias is rejected at the 1% confidence level), and therefore motivate the use of a method correcting for selection bias.

Table 3.3 shows estimation results of equation (3.9) by pooled 2SLS. The three potential issues identified in section 3.3.1 are now taken into account. Different specifications are presented in columns 1 to 4. In each case, *length* is instrumented by z, and the presence of a child under 3 is the exclusion variable in the selection

Table 3.3: Final step of the estimation

	Dependant variable *wage* (1)	(2)	(3)	(4)
length	-0.077***	-0.17***	-0.077***	-0.17***
	(0.020)	(0.023)	(0.020)	(0.024)
age	0.027***	0.035***	0.027***	0.035***
	(0.0055)	(0.0067)	(0.0056)	(0.0067)
*age*2	-0.021***	-0.027***	-0.021***	-0.026***
	(0.0068)	(0.0082)	(0.0068)	(0.0082)
nbchild	0.048***	0.067***	0.046***	0.068***
	(0.011)	(0.013)	(0.012)	(0.013)
$part - time$	-0.57***		-0.57***	
	(0.0057)		(0.0057)	
$temporal\ FE$	yes	yes	yes	yes
$regional\ FE$	yes	yes	yes	yes
Exclusion variables :				
Child under 3	yes	yes	yes	yes
Unemployment rate	yes	yes	no	no

Notes: Total sample consists of women in the EDP/DADS matched sample, who gave birth to a second-born child between 1986 and 2002. Wages are observed 204 177 times between 1984 and 2005. Coefficients on Mills ratios and time-averaged variables are not reported. The dependent variable *wage* is the log of the average daily wage earned in the corresponding year, in €2005.

equation. They differ in two dimensions: a part-time job indicator can be included in the set of covariates, and local unemployment rate can be used as another exclusion variable.

We find a negative causal impact of parental leave duration on later wages in all specifications, and these estimates are significant at the 1% confidence level. According to Table 3.3, each extra year of withdrawal decreases later wages by 7.7% when taking into account part-time jobs (columns 1 and 3). The loss is even bigger when this variable is not included in the regressors (-17%, columns 2 and 4).

3.6 Discussion

3.6.1 Causal impact magnitude

Our results suggest that parental leave duration has a negative causal impact on wages, on top of a possible wage penalty when women give birth[10]. According to Table 3.3, each extra year of withdrawal decreases later wages by 7% or 17% depending on the specification. A decrease of 7% per year corresponds to a 11% wage loss for the average withdrawal length in our sample (1.5 year). This effect is massive, and larger than one would have expected. At this stage, one must be very cautious when interpreting the magnitude of the effect. It represents the mean effect on wages mothers face during the remaining of their career. This does not necessarily mean that mothers' wage level decreases by 11% from their return to work until they retire. This decrease could be explained by a (time-constant) wage gap after the birth, or/and a lower wage growth rate. The cumulative effect of a 0.5% difference in wage growth rate during the 25 years between the birth of a child and the end of one's career would create a wage differential of 13% after 25 years.

Bayet (1997) can help us put these results into context. He studied wage differentials between French workers with interrupted or uninterrupted careers. By his definition a career is interrupted if the sums of spells out of the labor market during the career exceeds two years. He found that long breaks have a massive effect on wages. For example, focusing on female clerical workers with 25 years of total experience, and 10 years of tenure with their current employer (the average tenure in his sample), wages are 23% lower in case of interrupted careers. As Bayet (1997) does a ceteris paribus analysis, his results cannot be causally interpreted. However the magnitude of his coefficients is similar to ours.

3.6.2 Possible causal mechanisms

Several mechanisms could explain part of the causal effect of withdrawal on wages. Section 2.3.2 provides a detailed discussion on such causal paths, among which we now focus on the two main ones. First of all, experience is not taken into account

[10]See Felfe (2006).

in our model, but is known to contribute to wage growth. Bardaji et al. (2003) and Koubi (2003) estimate on French data that returns to one year of experience on wages are close to 2%. Ours results certainly incorporate this wage loss due to a lower experience.

Moreover, human capital may depreciate while out of work, which would result in lower wages. The magnitude of this loss would increase with the duration of the absence from work. This phenomenon is not specific to withdrawals after the birth of a child, and exists after most non-working periods. However Albrecht et al. (1999) claim that human capital deteriorates more during unemployment than parental leave. Once again, this contributes to the causal effect that we observe.

3.6.3 Structural and non-structural approaches

Chapters 2 and 3 are devoted to estimating the wage loss caused by parental leave duration. Both of them rely on the same dataset and the same instrument (the 1994 APE reform) to overcome the endogeneity of parental leave duration in the wage equation. However, they differ in their modeling strategy : the former uses a difference in difference (DiD) approach, whether the latter imposes more structure and specifies a selection equation.

Each approach has its own advantages and drawbacks. DiD are generally more robust to misspecification because they need very few structural hypotheses. Instead of having to specify how variables interact with each other, they only require that those unknown structural interactions do not change in a short period of time surrounding the 1994 reform. This *continuity* hypothesis is weaker than those required in a structural model, and seems credible in our context. However, DiD do not control for sample selection, since they do not take into account participation decisions at the individual level. This is not a problem, as long as women who give birth to a child just before the reform are similar to those whose delivery occurs just after the reform. If these two sets of women do not have the same distribution of characteristics, DiD estimations can be biased. On the other hand, the structural model presented in chapter 3 is not affected by this issue, since it finely controls for individual selection year after year. As usual, this is valid under the assuption that the selection equation is correctly specified.

These two complementary approaches give qualitatively similar results. Both suggest that parental leave duration has a negative causal impact on later wages and the magnitude of this impact is about the same (-10% in chapter 2, between -7% and -17% in chapter 3). This is a comforting hint that the structural assumptions made in section 3.3 do not seem to be too strong in our context.

An important difference between the two chapters is the estimations precision. Estimates are statistically significant at the 1% confidence level in all specifications of the structural model, whereas they reach the 5% confidence level only in one specification of the DiD approach. This may be due to the presence of covariates and individual fixed effects in the structural model, which decreases the part of unexplained variations in wages.

3.7 Conclusion

This chapter uses a structural model to test whether parental leave duration has a causal impact on later wages. Identification comes from a legislative change giving up to three years of benefits to mothers who decrease their labor supply after the birth of their second child. This monetary incentive has been having a massive impact on participation rate of eligible women from its implementation in 1994 until now (see section 3.2). This exogenous drop in participation allows us to identify the parameter of interest. Our model suggests that parental leave duration has a negative and significant causal impact on wages. Each extra year of withdrawal from the labor market decreases later wages by 7% or 17% depending on the specification.

This coefficient is close to those produced in a DiD estimation (chapter 2). This suggests that our structural model does not seem to be based on inappropriate assumptions.

3.8 Annex

3.8.1 Matching the EPD and DADS files

Figure 3.13 represents the share of EDP women also present in the EDP/DADS sample. This proportion is remarkably stable across birth cohorts, around 0.9. The selection between the initial DADS sample and the matched sample is plotted in Figure 3.14. The plain line represents the proportion of DADS observations corresponding to women present in the EDP. As expected, this proportion is roughly constant, close to 13% for even-numbered years of birth (4 days of birth selected out of 31 days in October). This ratio of 13% is also constant across years of presence in the DADS: Figure 3.15 pictures the proportion of observations sorted by year of presence in the DADS. The dotted line in Figure 3.14 represents a similar ratio, in terms of number of individuals in the DADS instead of number of observations. This curve is below the first one, between 10% and 13%. This is probably due to some wrong NIR remaining in the DADS sample[11].

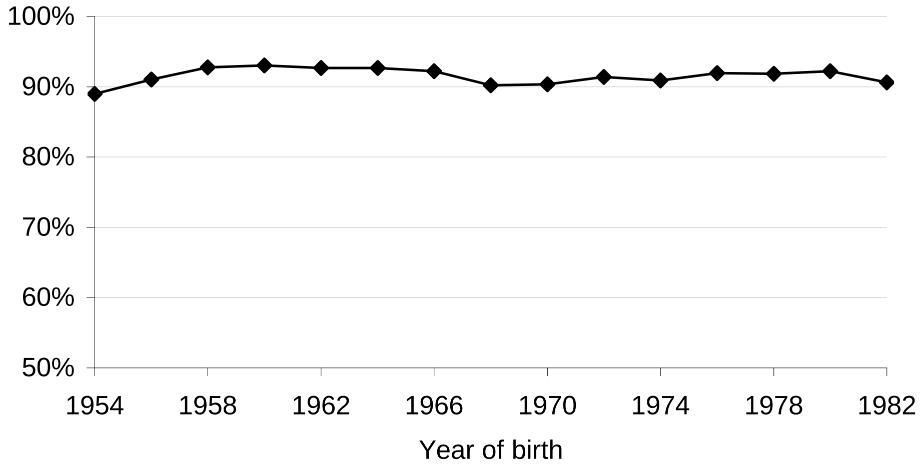
By construction, the EPD is representative of the French population, while the DADS may not be. To find out if the matching led to sample selection issues, we can check whether women's observable characteristics in the matched sample systematically differ from those of EDP. Figure 3.16 compares the number of children per woman between the EDP and the EDP/DADS matched sample. The proportion of women with no child is slightly higher in the matched sample (the gap is at most 29% vs. 25% for the 1970 cohort), whereas it is the opposite for mothers with two children. Overall, there is no huge difference.

3.8.2 Variance Covariance matrix

See Semykina and Wooldridge (2005).

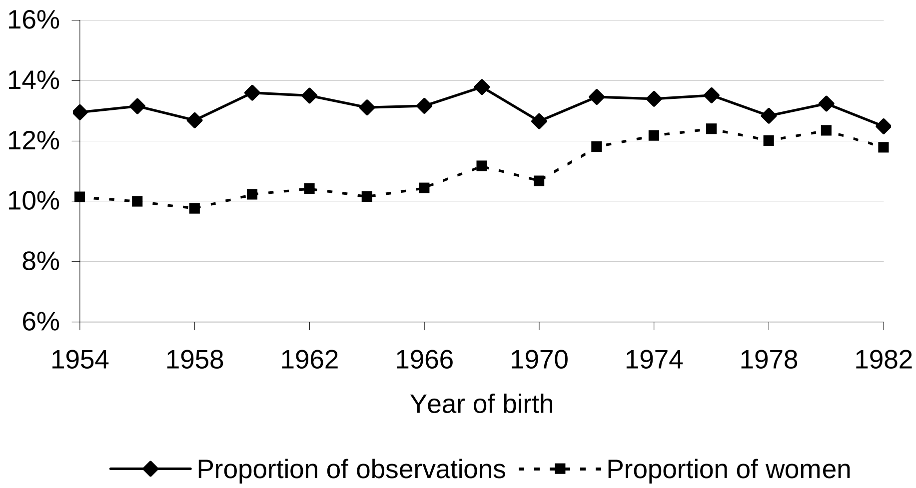
[11] An individual career generally consists in several observations in the DADS, since there is one observation per individual-firm-year. If the NIR is wrongly coded in one of these observations, it creates a new (fictitious) individual with a career reduced to only one observation. This could explain part of the difference between the two curves in Figure 3.14.

Figure 3.13: Proportion of EDP women present in the EDP/DADS sample



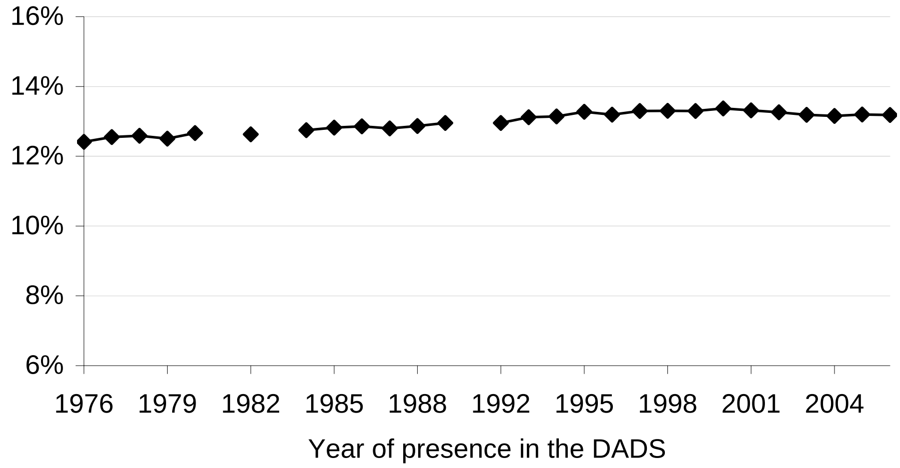
Notes: The curve represents the proportion of women in the EDP sample who are also present in the EDP/DADS matched sample, by birth cohort. Only even-numbered cohorts are plotted.

Figure 3.14: Proportion of DADS women and observations present in the EDP/DADS sample



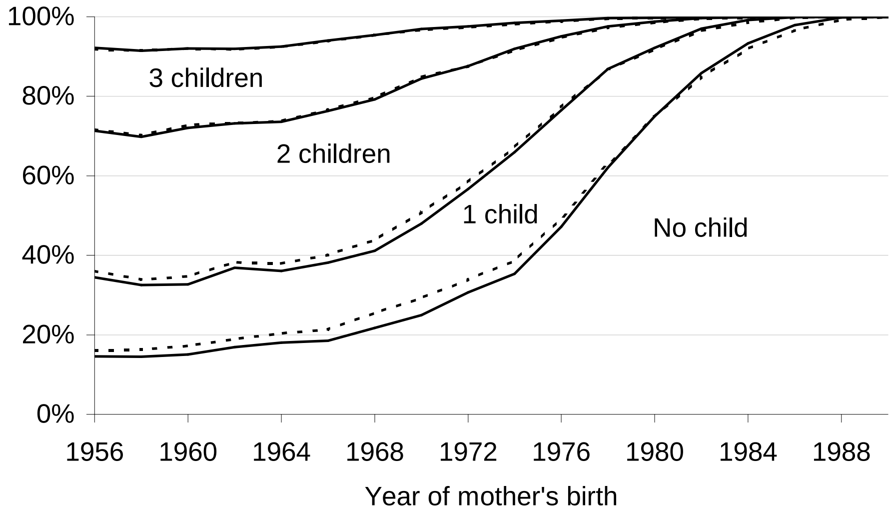
Notes: The plain line represents the proportion of DADS observations corresponding to women present in the EDP. The dotted line represents a similar ratio, in terms of number of individuals in the DADS instead of number of observations.

Figure 3.15: Proportion of DADS observations present in the EDP/DADS sample



Notes: This ratio of 13in the DADS: The curve represents for each year, the proportion of DADS observations that given year present in the EDP/DADS sample.

Figure 3.16: Number of children, comparison EDP vs. EDP/DADS



Notes: The curves represent the proportion of women with resp. 0, 1, 2, or 3 children, as a function of the woman year of birth. Dotted lines refer to the EDP, whereas plain lines refer to the matched sample EDP/DADS. Only even-numbered cohorts are plotted.

3.8.3 Testing for selection bias

Section 3.3.2 showed that a FE-2SLS estimation of equation (3.1) can give consistent estimates if selection $s_i = (s_{i1}, ..., s_{iT})$ is strictly exogenous (see equation (3.2)). Semykina and Wooldridge (2005) propose a procedure to test past or future selection bias. Adding the selection indicators $s_{it'}$ (for $t' \neq t$) to equation (3.1) and estimating the augmented equation by FE-2SLS should lead to non significant coefficients on the $s_{it'}$ if assumption (3.2) holds.

They also develop a method to test for contemporaneous selection bias in the FE-2SLS estimation of equation (3.1). This procedure (described in detail in Semykina and Wooldridge 2005) boils down to estimating

$$y_{it} = x_{it}\alpha + l_{it}\beta + c_i + \rho_t\widehat{\lambda_{it}} + \epsilon_{it}, \quad t = 1, \ldots, T \tag{3.12}$$

by FE-2SLS. If the ρ_t are jointly significant, then there is indeed contemporaneous selection. In that case, estimation of β by FE-2SLS is biased, and the procedure described in section 3.3 is required. Note that this method does not test for past or future selection bias. Therefore accepting the null hypothesis of no contemporaneous bias does not imply that FE-2SLS is consistent.

Table 3.4 shows that the Wald test of joint significance of the ρ_t rejects the null hypothesis of no contemporaneous selection bias at the 1% confidence level. Therefore the procedure described in section 3.3 is required to produce a consistent estimate of β.

Table 3.4: Testing for selection bias

	(1)	(2)
$mills_1$	-.2760801***	-.3381206***
	(.0446099)	(.0448014)
$mills_2$	-.2538073***	-.3065577***
	(.0434802)	(.0443483)
$mills_3$	-.2533781***	-.3134802***
	(.0450312)	(.0455259)
$mills_4$	-.2784707***	-.3676864***
	(.043574)	(.0448777)
$mills_5$	-.2903211***	-.4023698***
	(.0422251)	(.045117)
$mills_6$	-.2888452***	-.4055856***
	(.0434322)	(.0449973)
$mills_7$	-.4391497***	-.5795309***
	(.0490407)	(.0523085)
$mills_8$	-.5178682***	-.6619586***
	(.0507555)	(.0537436)
$mills_9$	-.4386951***	-.6033393***
	(.0591959)	(.0619414)
$mills_{10}$	-.3857685***	-.5626253***
	(.0577071)	(.0620861)
$mills_{11}$	-.4950894***	-.6997221***
	(.0600658)	(.0647957)
$mills_{12}$	-.4029094***	-.6006247***
	(.0599825)	(.0646308)
$mills_{13}$	-.4273978***	-.6426625***
	(.0608473)	(.0649924)
$mills_{14}$	-.4414156***	-.6398817***
	(.0631861)	(.067327)
$mills_{15}$	-.4152668***	-.6112592***
	(.0644735)	(.0689118)
$mills_{16}$	-.3297608***	-.506142***
	(.0622768)	(.0666462)
$mills_{17}$	-.4879121***	-.6823519***
	(.0699378)	(.0742196)
$mills_{18}$	-.2889029***	-.4675414***
	(.0568053)	(.0606741)
$mills_{19}$	-.2865194***	-.4610958***
	(.0614204)	(.0655767)
$mills_{20}$	-.2656346***	-.4817236***
	(.0600736)	(.0643062)
$mills_{21}$	-.294866***	-.524047***
	(.0691163)	(.0727939)
Joint Wald test on Mills ratios	F(21, 17388) = 7.45 Prob > F = 0.0000	F(21, 17388) = 10.83 Prob > F = 0.0000

Notes: Total sample consists of women in the EDP/DADS matched sample, who gave birth to a second-born child between 1986 and 2002. Coefficients from the FE-2SLS estimation of equation 3.12 (second step). A part-time job indicator is included in the set of covariates (column 1) or not (column 2). Mills ratios are numbered chronologically ($mills_1$ for 1984, up to $mills_{21}$ for 2005, 1990 is missing).

Deuxième partie

Education et santé

Chapter 4

Does compulsory education lower mortality?

Sommaire

This chapter was written with Valérie Albouy (Insee), and published as: Albouy, V. and Lequien, L. (2009), 'Does compulsory education lower mortality?', *Journal of Health Economics* **28**(1), 155–168.

Until recently, literature on return to education was mostly focused on monetary gains stemming from a higher level of education. Researchers tried to identify to what extent higher wages earned by more educated people were due to an improvement in human capital. A major extension consists in exploring non-monetary returns to education. Besides bettering standard of living, education has an influence on many aspects of daily life. It modifies social behaviors as diverse as fertility, spending profile, wearing a seat belt or voting participation (Milligan et al. 2004). Our purpose is to determine whether education increases one's health capital. Hence we want to test if a given individual, when provided with a higher level of education, has a better subsequent health than in the absence of extra education.

A strong empirical correlation between education and health is now well established, and the debate among economists currently lies on possible mechanisms explaining that correlation. Two types of arguments are generally put forward. In the first place, there could be a third variable influencing both level of education and health status, and thus indirectly creating a link between them. Familial background and times preferences are likely candidates for creating an artificial correlation between education and health. On the other hand, several mechanisms could create causal links between education and health. Grossman (1972) proposed two causal paths linking education to health. Firstly, educated people may have a greater productivity, and thus could produce a greater amount of health from a given quantity of inputs. This argument is usually put forward to explain why education may have a causal impact on wages: educated individuals are more productive, which results in higher wages in the labor market. In the same way, education could favorably affect productivity in nonmarket activities like health production (Michael 1973). This could stem from a better use of available information made by educated people. Spandorfer et al. (1995) observed that patients with low literacy scores were more likely to have poor comprehension of their emergency department discharge instructions. The more educated also seem to better comply with complex medical treatments (Goldman and Smith 2002). Moreover, Grossman (1972) assumed that educated people may be able to allocate available inputs to produce health more efficiently than less educated people. They may indeed have a better access to information, and therefore benefit more from recent improvements in medical techniques (Glied and Lleras-Muney 2003). De Walque (2004) claimed that smoking behaviors reacted earlier and more intensely among the educated after the dangers related to tobacco consumption began to be publicized. They may also be more inclined to build and follow a balanced diet. This strand of the literature suggests that education might improve general skills relevant to health related behaviors, besides the specific knowledge acquired at school.

Apart from these mechanisms, a channel through which causality may run is via peer effects. The more one studies, the more one is in contact with educated people. This social segregation begins during school years, and then goes on in the workplace. If indeed educated people adopt healthy behaviors more frequently, being surrounded by educated friends and colleagues could increase the likelihood of behaving healthily as well (Katz et al. 2001). Eventually highly educated people have greater income, and thus can more easily afford preventive care, costly medical treatments, or to live in less polluted areas (Jusot 2003). Even if the relationship between education and health is not a direct one, since it goes through income, it is still a causal link: increasing one's education level will lead to higher income, and this will have a positive impact on one's health.

We investigate the existence of a causal relationship from education to health, using two successive increases in compulsory education laws in France during the XX^{th} century. Minimal school leaving age was first increased from 13 to the age of 14 for cohorts born after 1923 (Zay reform), and was extended by two more years for individuals born after 1953 (Berthoin reform). A relevant framework to exploit this kind of changes is the regression discontinuity approach, based on the fact that these reforms created two discontinuities in school attainment. The next section is a survey of the growing literature on the link between education and health. Then we detail the econometric framework we implemented, along with our nonparametric and semiparametric estimation strategies. Section three describes the dataset, while our results are presented in section four. The last section concludes with some suggestions for further research.

4.1 Previous studies

When studying the impact of education on health, the first methodological problem arises when one wants to measure people's health: health is not a one-dimensional variable. Defined as "state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity" by the World Health Organization in its 1946 Constitution, health is a fairly general concept. This lack of precision results in numerous health measures used in the literature on the link between education and health[1] . Regarding morbidity, Berger and Leigh (1989) studied blood pressure, Groot and Maassen van den Brink (2007) focused on illnesses prevalence, Doyle et al. (2005) used chronic conditions. Body mass index is the health proxy in Spasojevic (2003), Chou et al. (2004), Arendt (2005). Mortality is studied by Lleras-Muney (2005), Elo and Preston (1996) on American data, van Oers (2003) on Dutch data, and Bopp and Minder (2003) on Swiss data. As for subjective measures of health, Adams (2002), Arendt (2005), Spasojevic (2003), Oreopoulos (2006), and Doyle et al. (2005) used self-rated health when they evaluated the impact of education on health. The latter can also be described with mobility limitations or difficulties to carry out daily activities (Adams 2002, Arkes 2003, Oreopoulos 2006, Berger and Leigh 1989). Eventually Kenkel (1991), De Walque (2003, 2004, 2007), Arendt (2005) and Kenkel et al. (2006) studied the influence of education on adoption or rejection of risky behaviors. It appears that all these papers find a strong correlation between the level of education and the measure of health they selected. This result still holds when income is controlled for.

To investigate whether causality between education and health may explain (at least part of) that correlation, most authors implement an instrumental variable strategy. These instruments can either be institutional parameters, macroeconomic variables, or instruments specific to education issues (see Table 4.1). It appears that the very existence of a causal impact of education on health is not clearly established. Some authors claim to find causality with all the health measures they use (Oreopoulos (2006) with physical limitations and self-rated health), others only

[1] See Grossman (2004) for a comprehensive survey.

with some of them (Adams (2002) with some physical limitations but not all). As for Groot and Maassen van den Brink (2007), they find a significant causality for men but not for women. Eventually several studies (Arendt 2005, Auld and Sidhu 2005) do not show any significant causal effect, despite testing the causality assumption on several health measures. The hypothesis that education would have an influence only on specific health dimensions could explain this lack of consensus, but it seems not to be the case. There are indeed contradictory results for several given measures of health (e.g. Arkes (2003) and Oreopoulos (2006) on physical limitations).

The use of different instrumental variables does not seem to be the reason for such diverse results either. For instance, among the seven papers listed in Table 4.1 which use changes in compulsory education laws as an instrument, four report a significant causal impact, two do not, and the last one has mixed results depending on the health measure considered. A possible explanation could be that they have various health measures, and that this type of instrument can only reveal differences in health on particular health characteristics. This assumption does not hold since two papers (Oreopoulos (2006) on UK data and Arendt (2005) on a Danish sample) end on opposite conclusions with the same health measure (self-rated health) and the same instrument. Discrepancy could also stem from fundamental differences between countries, since authors of the seven aforementioned studies work on datasets covering four countries (USA, Sweden, United Kingdom, and Denmark). Yet it is hard to conceive causal mechanisms appropriate only to some of these countries. On the other hand, the power of the instrument can differ among countries. Depending on whether the rise in mandatory school leaving age was indeed a constraint for a large part of the population, the instrument will be more or less correlated with education. Eventually precision of estimations generally depends on the sample size. This could partly explain why some studies bring out a significant causality between education and health, while some others do not.

Table 4.1: Causality between education and health

Authors	Health measure	Source, country, year, sample size	Instrument	Significant causality	Remarks
Berger and Leigh (1989)	Blood pressure, activity limitations between 20 and 40 years old	NES, USA, 1970, 13 500; NLS, USA, 1966-1971, 3 600	Per-capita State expenditures on education, per-capita disposable income, IQ test	Yes	
Kenkel (1991)	Smoking, drinking, lack of physical activity	NHIS, USA, 1985, 33 000	Access to anti-alcohol and anti-tobacco public campaign	Yes	
Adams (2002)	SRH, activity limitations between 51 and 61 years old	HRS, USA, 1992, 24 000	Change in compulsory education laws, uarter of birth	Yes /no	Causality only for some limitations
Spasojevic (2003)	Health index, BMI at 50 years old	Sweden	Change in compulsory education laws	Yes	One fifth of education effect on health is an income effect

Continues on next page. . .

Table 4.1 – Continued

Authors	Health measure	Source, country, year, sample size	Instrument	Significant causality	Remarks
De Walque (2003)	Smoking	IFLS, Indonesia, 1993, 3 000	Date of school construction	Yes	
Glied and Lleras-Muney (2003)	Mortality after 70 years old	SEER, USA, 1973/1993, 600 000; CMF, USA, 1960/1990, 250 000	Change in compulsory education laws	Yes	
Arkes (2003)	Activity limitations, work limitations, require personal care, between 47 and 56	Census, USA, 1990, 400 000	Local unemployment rate	Yes /no	Causality on work limitations and personal care, but not on activity limitations
De Walque (2004)	Smoking	NHIS, USA, 1978-2000, 370 000	Access to anti-alcohol and anti-tobacco public campaign	Yes	

Continues on next page. . .

Table 4.1 – Continued

Authors	Health measure	Source, country, year, sample size	Instrument	Significant causality	Remarks
Doyle et al. (2005)	SRH, chronic health condition for children aged 8	HSE, UK, 1997-2002, 7 000 children	Change in compulsory education laws, grand-parents smoke	No	No causal effect of income either
Arendt (2005)	SRH, BMI, smoking	WECS, Denmark, 1990/1995, 3 300	Change in compulsory education laws	No	
Lleras-Muney (2005)	Ten-year mortality, around 50 years old	Census, USA, 1960-1980, 800 000	Change in compulsory education laws	Yes /no	Regression discontinuity estimates are not significant
Auld and Sidhu (2005)	Work limitations, between 36 and 43	NLSY79, USA, 2000, 6 400	Parents level of education, local unemployment rate	No	
Kenkel et al. (2006)	Smoking, BMI between 34 and 41	NLSY79, USA, 1998, 6 500	5 education policy variables at State level (expenditures, etc.)	Yes /no	Causality only on smoking for men, when education is measured with "having completed high school or not"

Continues on next page...

Table 4.1 – Continued

Authors	Health measure	Source, country, year, sample size	Instrument	Significant causality	Remarks
Oreopoulos (2006)	SRH, activity limitations between 25 and 84 years old	GHS, UK, 1983-1998, 66 185; Census, USA, 1990/2000, 1 000 000	Change in compulsory education laws	Yes	Implement an IV/RD framework
Cipollone et al. (2006)	Mortality between 25 and 35	Census, Italy, 1981, 1981/1991	Exemption from military service	Yes	
Groot and Maassen van den Brink (2007)	SRH, chronic condition	DSCPB, Netherlands, 1999, 13 500	Father has a managerial job, number of workers supervised, mother works	Yes /no	No causality for women, just significant for men
De Walque (2007)	Smoking	NHIS, USA, 1983-1995, 80 000	Vietnam era draft lottery	Yes /no	Causality on probability of smoking, but not on probability of quitting in all specifications

Note: SRH stands for self-rated health, BMI is the body mass index.

To the best of our knowledge, mortality is currently an exception among health measures, since no paper has challenged the causality that Glied and Lleras-Muney (2003), Lleras-Muney (2005) and Cipollone et al. (2006) have found. Moreover, the causal link between education and health has not been studied on French data yet. Hence the goal of this paper is twofold: we wish to add the French case to the existing literature, and test the existence of a causal mechanism between education and mortality on French data. We put our work in the specific setting of regression discontinuities, as we believe it is the more appropriate framework for an analysis using school attainment legislation.

4.2 Methodology

In the literature on program evaluation, the probability of receiving a treatment usually depends (at least partly) on the value of a continuous variable. The key feature of regression discontinuity designs is that for a given cutoff value of this variable, the probability of having been treated is discontinuous. The idea to exploit discontinuities in treatment assignment probabilities was first introduced by Thistlethwaite and Campbell (1960). More recent works include the evaluation of the impact of class size on student's achievement (Angrist and Lavy 1999) or the influence of financial aid amount a college offers on students' decision to enroll in this college (van der Klaauw 2002). Hahn et al. (2001) set up the theoretical framework for regression discontinuity designs. Using their notations, let x_i be the binary treatment indicator: $x_i = 1$ when individual i follows the treatment and $x_i = 0$ otherwise. There are two possible outcomes: y_{0i} and y_{1i}. For each individual i, y_{0i} is health status if individual i has not been treated, y_{1i} if he has received the treatment. For individual i, the treatment effect equals $y_{1i} - y_{0i}$. As usual in evaluation studies, either y_{0i} or y_{1i} is observed for a given individual i, but not both of them. Let y_i be the observed outcome. y_i can be expressed as

$$y_i = \alpha_i + \beta_i x_i$$

where $\alpha_i = y_{0i}$ and $\beta_i = y_{1i} - y_{0i}$. β_i is the treatment effect for individual i.

It may be heterogeneous among individuals. The probability of being treated is a function of an observed variable z_i. The basic feature in a regression discontinuity design is the following:

$$(H_0) \qquad \exists z_0 | \lim_{z \to z_0^-} E(x_i|z_i = z) \neq \lim_{z \to z_0^+} E(x_i|z_i = z)$$

(H_0) means that the probability of being treated, regarded as a function of z, is not continuous in at least one value z_0 of z. In our context, the treatment consists in attending school at least a given number of years. zi is the year of birth, and the cutoff values are the years education reforms came into force. Raising mandatory school leaving age induced a fraction of the population to go on studying until they reached the new legal limit, which increased the proportion of treated. Obviously some individuals were already following the treatment before these reforms were implemented. This means that selection into treatment is not based solely on z_i, but also on other (un)observed variables. This case is usually referred to as a fuzzy design.

In case of a constant treatment effect, the minimal identification condition is

$$(H_1) \qquad E(\alpha_i|z_i = z) \text{ as a function of } z \text{ is continuous at } z_0$$

meaning that in absence of treatment, individuals born just before z_0 and those born just after would have similar average outcomes. It implies that individuals whose z_i is close to z_0 are on average essentially the same as far as health is concerned, even in dimensions which are not observed by econometricians. Under assumptions (H_0) and (H_1), the constant treatment effect β is nonparametrically identified (Hahn et al. 2001):

$$\beta = \frac{y^+ - y^-}{x^+ - x^-}$$

$$\text{where} \quad \begin{cases} y^+ = \lim_{z \to z_0^+} E(y_i|z_i = z) & y^- = \lim_{z \to z_0^-} E(y_i|z_i = z) \\ x^+ = \lim_{z \to z_0^+} E(x_i|z_i = z) & x^- = \lim_{z \to z_0^-} E(x_i|z_i = z) \end{cases}$$

Since (H_1) guarantees that the only parameter relevant to health distinguishing individuals above and below the cutoff value is education, differences in health between these two populations can be causally attributed to the differential in education. The treatment effect is equal to the difference in health, divided by the difference in treatment participation.

If the treatment effect is allowed to be heterogeneous among the population, two further assumptions are needed to identify a causal effect (Hahn et al. 2001):

(H_2) $(\beta_i, x_i(z))$ are jointly independent of z_i in a neighborhood of z_0

(H_3) $\exists\, \epsilon > 0 \mid \forall e \in [0;\epsilon],\ x_i(z_0 + e) \geq x_i(z_0 - e)$

(H_2) is more flexible than the classical assumption of no selection into treatment. Self-selection is indeed allowed; the only requirement is that this selection would have followed the same rules just before and just after the cutoff point z_0 in the absence of reform. In particular, individuals can select into treatment according to the treatment effect they expect. In our context it is unlikely that individuals decide to attend school depending on the gain they expect to their health state, because this return is hard to quantify. Yet, they may make their education choices according to their return to education in terms of wages. If the link between education and health passes through an income effect, such a behavior rules out any identification strategy based on the absence of selection into treatment. Monotonicity assumption (H_3) means that individuals around z_0 must all react to the assignment to treatment in a monotone (increasing) way. In other words, noboby can be deterred from following the treatment by being assigned to the treatment. Using Angrist et al.'s (1996) terminology, (H_3) assumes the absence of defiers: individuals who follow the treatment when they are not assigned to, and do not take it when they are assigned to. If the treatment consists in studying at least until 16 years old, a raise in mandatory education from 14 to 16 should not induce someone to decrease her/his school leaving age from 16 to 15. Although its validity can not be empirically tested, (H_3) is likely to hold in our context. Under (H_0), (H_1), (H_2) and (H_3) the average treatment effect for compliers can be nonparametrically estimated:

$$\lim_{e \to 0^+} E(\beta_i | x_i(z_0 + e) - x_i(z_0 - e) = 1) = \frac{y^+ - y^-}{x^+ - x^-}$$

Compliers are those who were induced to follow the treatment by the reform. Here, compliers are those who were induced by the legislative change to extend their schooling above the new legal age. From a public policy perspective, compliers are the relevant subsample to evaluate the impact of a reform since they are the ones directly affected by the policy change.

Regression discontinuity designs provide an estimation strategy which relies solely on individuals whose z_i is very close to the discontinuity point z_0. The principle is that z is a "local" instrumental variable: the discontinuity guarantees a strong correlation between z_i and x_i; reasoning at the limit ensures that the effect of z_i on y_i passes only through its effect on x_i. But it might be the case that this restricted sample is too small to allow the identification of a significant causal effect. Unfortunately, adding to the sample individuals less close to z_0 is not sufficient to overcome that problem, because z may not be a valid instrument on the extended sample. For instance, in the context of educational returns on health, birth cohort z_i is not a valid instrument since this exclusion condition may not hold due to a "generation effect": it is likely that scientific knowledge, better preventive care or better personal hygiene induced an improvement in average health state at the age of 50 through generations. It is however possible to use information from a larger part of the sample under additional assumptions (van der Klaauw 2002). In a more classical setting, the economic model can be written with a standard endogenous dummy variable model:

$$y_i = \alpha + \beta x_i + u_i$$

with β the treatment effect (assumed to be constant only for presentation purposes) and x_i the dummy variable indicating participation into treatment. Formally, the "generation effect" means that u_i and z_i are correlated in the extended sample. However z_i may still be used as an instrumental variable if it is adequately controlled for. Indeed, the solution is to introduce an estimation of $\mathbb{E}(u_i|z_i)$ as covariate in the equation above. Assuming that $\mathbb{E}(u_i|z_i)$ is a continuous function of z_i at $z = z_0$ enables us to identify the variation in y_i attributed to the discontinuity of x_i at z_0 and to infer the local treatment effect. Doing so, it is important to allow the functional form of $\mathbb{E}(u_i|z_i)$ to be as flexible as possible, since any misspecification may lead to biased estimations. $\mathbb{E}(u_i|z_i)$ is usually parameterized as a polynomial function of z_i. Thus, the model will take the following form:

$$y_i = \alpha + \beta x_i + k(z_i) + w_i$$

where $k(z_i)$ represents $\mathbb{E}(u_i|z_i)$. Note that by construction the error term w_i is not correlated with z_i and that under (H_1) it is continuous at $z = z_0$. From now on, the problem of endogeneity of x_i, due to a possible selection into treatment, can be solved by a classical two-step instrumental method using z_i as an instrument. Identification of β stems from the fact that x is the only variable which depends discontinuously on z at $z = z_0$.

If we relax the assumption of constant treatment effect, and allow β_i to vary across individuals, the previous methodology remains valid under assumptions (H_1), (H_2) and (H_3). In that case, $k(z_i)$ does not model $\mathbb{E}(u_i|z_i)$ any more but $\mathbb{E}(u_i|z_i) + \{\mathbb{E}(\beta_i|z_i) - \mathbb{E}(\beta_i|z_0)\}\mathbb{E}(x_i|z_i)$ and the procedure gives an estimation of $\mathbb{E}(\beta_i|z_0)$ (see van der Klaauw 2002).

The last methodological point is about the definition of treatment. Literature on policy evaluation has widely used binary treatments, even when treatment is not homogeneous across individuals (e.g. Rosen and Willis 1979, for return to education on wages). But using a binary treatment does not allow the estimation of the marginal effect of schooling on health. The framework depicted above can be extended to cases where treatment intensity can vary. Angrist and Imbens (1995) showed that the Wald estimator overestimates the marginal treatment effect when a treatment with variable intensity is specified with a dummy equal to one when intensity x is greater or equal than a given threshold. However there is no overestimation if the sole impact of the reform is to induce some individuals to extend their schooling by exactly one unity of treatment. Therefore we decided to use a treatment with intensity for the two reforms, and also to provide results with a binary treatment for the Zay reform which extended mandatory schooling by only one year.

4.3 Data

Our empirical work is based on the Echantillon Démographique Permanent (EDP thereafter). The EDP is a longitudinal dataset containing information on a one per cent sample of the population living in France. It was built from the 1968 national census, by selecting all individuals living in France and born on the four first days of October. This longitudinal dataset has been growing each year with the addition of newborns of these four days of October. For all these individuals, data from the 1968, 1975, 1982, 1990 and 1999 national censuses are gathered, as well as information from the register of births, marriages and deaths from 1968 until 2005. We excluded from our sample those born abroad, because the matching process between census files and datasets containing death certificates is less reliable for these individuals. Answering to the national census is mandatory for everyone living in France. Moreover, death certificates are extracted from an exhaustive dataset at the country level. Hence the only possible source of attrition between the creation of the EDP in 1968 and 2005 may stem from international migration. More precisely, individuals present at least in one census wave may die later while abroad. In such a case, death certificates are likely to be missing. However, two reasons lead us to believe that attrition may not be a major issue in our context. Firstly, deaths occuring while abroad certainly account for a small proportion of the total number of deaths. Moreover there is no obvious reason to believe that migration might be correlated to the fact of being born right before or after the reforms, and therefore this possible attrition can be seen as random within our framewwork.

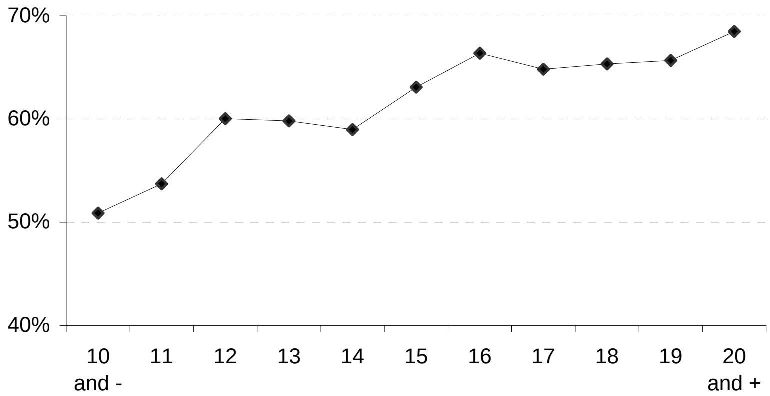
Censuses provide us with people's school leaving age. Our measure of health is mortality. We computed survival rates at a given age, taking advantage of the longitudinal structure of the EDP. Most papers relying on changes in school leaving age use cross-sectional data and have to take into account that people who studied under the pre-reform laws are older than those who studied after the reform. As a result, there might be differences in health between those two populations just because health is not measured at the same age for everyone. Authors generally add age as a covariate, in order to control for this age effect. But estimations will be biased if the age specification is not correct (in most cases if the effect of age on health is not linear). We are able to avoid this pitfall by measuring health status at the same age for everyone. For instance, as we have information on death up to 2005, survival rates are computed at the age of 80 (resp. 50) when we consider 1920 to 1925 (resp. 1950 and 1955) cohorts.

We selected individuals born at most three years before or after Zay and Berthoin reforms[2]. It represents 35 828 persons born between 1920 and 1925, and 47 337 persons born between 1950 and 1955 respectively. Figures 4.1 and 4.2 show the correlation between school leaving age and survival rate. A clear pattern is visible: the probability of being still alive at 50 or at 80 is growing with school leaving age. As expected, survival variability is larger at 80 than at 50, and survival rates at 50 are beyond 90

Let's now focus on level of education. Figures 4.3 and 4.5 represent mean school leaving age by birth cohort. We notice the well-known positive trend, accounting for the general increase in education during the XXth century. The mean increase in education between 1920 and 1960 is around one month per year. Vertical lines indicate the dates Zay and Berthoin reforms came into effect. A sudden break is visible in 1953 (Figure 4.3): mean school leaving age increased by 4 months between 1952 and 1953. It comes from the change in compulsory education laws which extended mandatory schooling from 14 to 16 for those born after January 1st 1953. In theory, every individual born from 1953 on should have been to school until

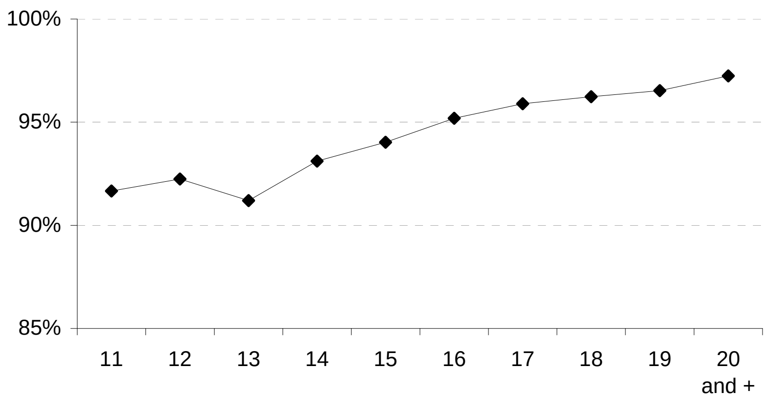
[2]By construction of the EDP, individuals born before 1968 have to be still alive in 1968 to appear in the EDP. To take into account this sample selection, we excluded from our Zay sub-sample those born between 1920 and 1925 who died before turning 48. We applied the same method on the Berthoin sub-sample, and kept only those still alive at 18. Therefore our health measures must be understood as *survival rate at 80 (resp. 50) given that one is still alive at 48 (resp. 18).*

Figure 4.1: School leaving age and survival rate at 80 years old, Zay reform



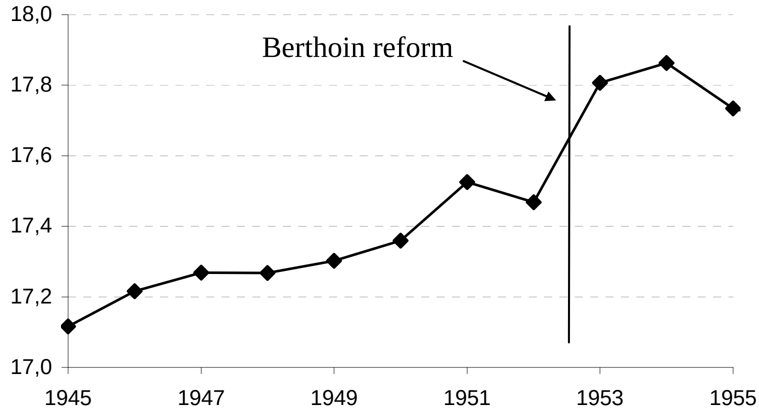
Scope: 35 828 individuals born between 1920 and 1925.

Figure 4.2: School leaving age and survival rate at 50 years old, Berthoin reform



Scope: 47 337 individuals born between 1950 and 1955.

Figure 4.3: Mean school leaving age by birth cohort, Berthoin reform

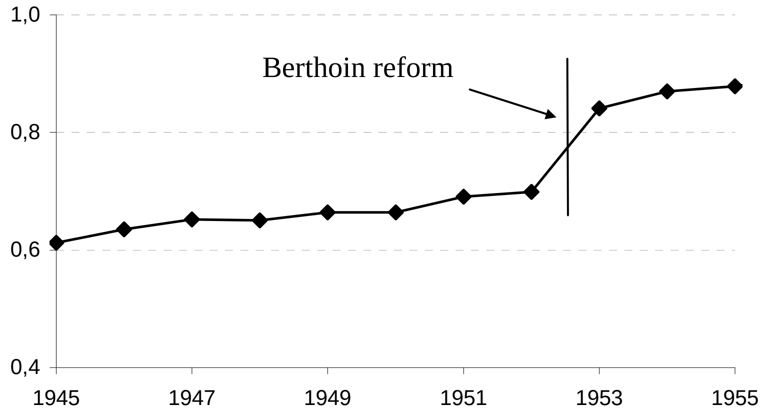


Scope: Individuals born between 1945 and 1955.

the age of 16. However this is not what we observe in our data. The proportion of people leaving school after 16 does increase sharply between 1952 and 1953 (see Figure 4.4), but there are still 15% of generations after 1953 who give up school before 16. This may be due to exemption clauses for work purpose. Yet this fact does not jeopardize our estimation strategy, since identification only requires that the institutional change induced a discontinuity in the average school leaving age.

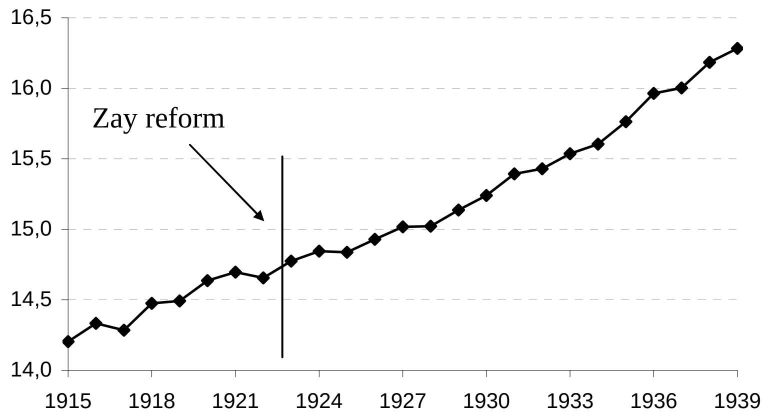
The discontinuity induced by the Zay reform is less obvious (see Figures 4.5 and 4.6). Its magnitude is logically less important than in the Berthoin case, because Zay extended compulsory education by only one year, whereas Berthoin raised mandatory education by two years. Moreover, two historical reasons tend to smooth the impact of Zay reform on education. Lower secondary education gradually became free in France between 1928 and 1933 (Prost 1968), which led to a dramatic increase in the number of students in secondary schools in the beginning of the 30ies. When Zay reform came into force in 1936, its impact on education was lessen because of that previous increase in education. The second reason is World War II, which created a drop in education level for those who left school after 1939. It is worth noticing that the war did not have any impact until the 1926 generation on the share of each generation leaving school after turning 14. So this education indicator

Figure 4.4: Proportion of each generation still studying at the age of 16, Berthoin reform



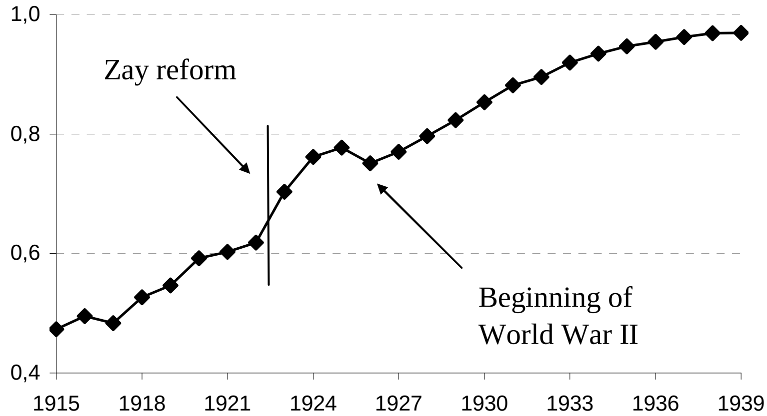
Scope: Individuals born between 1945 and 1955.

Figure 4.5: Mean school leaving age by birth cohort, Zay reform



Scope: Individuals born between 1915 and 1939.

Figure 4.6: Proportion of each generation still studying at the age of 14, Zay reform



Scope: Individuals born between 1915 and 1939.

was not affected by the war for cohorts in our estimation sample (between 1920 and 1925). On the contrary, school leaving age was disrupted before 1925 on the high tail of the distribution[3]. As we wish to draw causal conclusions out of variations in education, it is essential to know precisely what are the sources of variation in education variables and how they affect these variables. Added to the fact that the interpretation of return to education in the Zay case does not depend on the treatment definition (binary or with variable intensity), it strengthens the use of a binary treatment with the Zay reform.

4.4 Results

Before investigating the very nature of the relationship between education and health, we first have to check whether there is indeed a significant correlation in our data. Figures 4.1 and 4.2 suggest that a strong positive link exists between school leaving age and survival at 50 or 80 years old. We regressed health status y_i

[3]The decrease in education induced by the war was focused mainly on cohorts born after the implementation of the Zay reform. Hence the war may have reduced the size of the discontinuity in school leaving age created by the Zay reform.

Table 4.2: Regression of survival rate on school leaving age

	Zay reform y = survival at 80	Berthoin reform y = survival at 50
Intercept	-0.19***	0.83***
	(0.036)	(0.06)
Education	0.028***	0.049***
	(0.0024)	(0.0035)

Standard errors in parenthesis. Scope: Individuals born between 1920 and 1925 for the Zay reform, between 1950 and 1955 for the Berthoin reform. Sample sizes are respectively 35 828 and 47 337.
*** Significant coefficient at the 1% confidence level.

of individual i (a dummy variable equal to 1 if i is still alive at 50 (resp. 80), and 0 otherwise) on his school leaving age x_i. Result of these probit estimations confirms that education is indeed correlated with survival rates at 50 and 80 years old (Table 4.2). This correlation does not necessarily mean that education has a causal impact on health. Omitted variables or a mix of unobserved variables and causal mechanisms could also create that correlation. To test the existence of causality, we now implement a regression discontinuity design.

Under assumptions (H_0), (H_1), (H_2) and (H_3), the Wald quotient estimates the local average treatment effect for the compliers. Empirical counterparts of the four limits in the Wald estimator can be estimated by restricting the estimation sample to individuals just above and below the discontinuity z_0. As the Berthoin reform came into force for those born after January 1st 1953, we used the 1952 cohort to compute left limits y^- et x^- and the 1953 cohort to compute right limits y^+ et x^+. We also provided two alternative specifications, by gradually widening the number of cohorts around 1953 up to 6 cohorts. Table 4.3 shows the results of such estimations for the Berthoin reform, while Table 4.4 is devoted to the Zay reform. Health outcome y is a dummy equal to 1 if the individual is still alive at a given age. The treatment variable x is the school leaving age. Following last section discussion, we also used a binary treatment (whether or not studied after 14 years old) in the Zay case. From Table 4.3, one extra year of schooling increases by 0.013

Table 4.3: Wald estimators for the Berthoin reform

Generations	Wald estimator x = school leaving age
1952/1953, y= survival at 52	0.013
	(0.011)
1951/1954, y= survival at 51	0.0072
	(0.0072)
1950/1955, y= survival at 50	0.0033
	(0.0061)

Standard errors in parenthesis, estimated by bootstrap. Sample sizes are respectively 15 518, 31 305, and 47 337.

the probability of being still alive at 52 years old. As expected, education indeed lowers mortality, but this treatment effect is not statistically different from 0 at the 10% confidence level. When we gradually widen the estimation window, neither the 1951-1954 nor the 1950-1955 Wald estimators are significant at the 10% confidence level.

Similar results on the Zay reform are presented in Table 4.4. Once again, all Wald estimators are positive but not statistically significant. Survival is measured around the age of 80, which seems to be old enough to observe differences in mortality. Yet it is possible that the Zay reform had too light an impact on education to significantly affect subsequent survival rates: it increased mandatory education by only one year, and both free secondary school and World War II reduced its impact on education.

The previous estimations were computed on samples containing between 2 and 6 births cohorts. It is important to keep in mind that regression discontinuity estimates are unbiased only if the populations below and above the cutoff point are identical in all dimensions relevant to health, except education. The validity of this continuity assumption becomes less and less likely as we widen the number of generations in our estimations, which is why we limited to 6 the maximum number of cohorts[4]. As described in section 4.2, the proper way to extend the number

[4]Distribution of observed variables such as sex and socio-professional group are quite similar

Table 4.4: Wald estimators for the Zay reform

Generations	Wald estimator x = school leaving age	Wald estimator x = still study at 14
1922/1923, y= survival at 82	0.063	0.10
	(0.62)	(0.11)
1921/1924, y= survival at 81	0.052	0.049
	(0.067)	(0.053)
1920/1925, y= survival at 80	0.045	0.048
	(0.035)	(0.036)

Standard errors in parenthesis, estimated by bootstrap. Sample sizes are respectively 11 586, 23 519, and 35 828.

of generations in a regression discontinuity framework is to implement a two-step procedure. The first step equation is:

$$E(x_i|z_i) = a + bz_i + c(z_i - z_0)\mathbb{1}_{z_i \geq z_0} + \delta\mathbb{1}_{z_i \geq z_0}$$

where δ explicitly captures the discontinuity in education caused by the reform (z_0 equals 1923 for the Zay reform and 1953 for the Berthoin reform). The control function is linear in the year of birth z, its slope is allowed to be different before and after the reform date z_0.

Average school leaving age increases by 0.06 year each year between 1950 and 1952 (column 3 of Table 4.5). The magnitude of this coefficient is in line with the long-term temporal trend observed between 1920 and 1960. The coming into force of the Berthoin reform in 1953 causes a highly significant 0.28 year rise in education. This sudden increase seems to be absorbed during the following years as yearly growth rate becomes negative between 1953 and 1955. It is indeed a temporary phase because level of education increases at roughly the same rate in the fifteen years preceding the Berthoin reform and the fifteen years following it. The Zay reform created no clear discontinuity in average school leaving age. This

in the six birth cohorts, which is a good sign that extending to six years does not invalidate the regression discontinuity approach.

Table 4.5: First stage estimations

	Zay reform		Berthoin reform
	Binary treatment	Variable intensity treatment	Variable intensity treatment
a	0.33***	14.51***	14.58***
	(0.009)	(0.53)	(1.25)
b	0.013***	0.0071	0.056**
	(0.004)	(0.025)	(0.025)
c	0.023***	0.028	-0.098***
	(0.006)	(0.036)	(0.035)
δ	0.080***	0.11	0.28***
	(0.011)	(0.064)	(0.063)

Standard errors in parenthesis. In column 1 (binary treatment), the dependent variable is a dummy equal to one if individual was still attending school at the age of 14. School leaving age is the dependent variable in case of a treatment with variable intensity. We substracted 1900 to the year of birth. The Berthoin (resp. Zay) sample contains 47 337 (resp. 35 828) individuals born between 1950 and 1955 (resp. 1920 and 1925).
** Significant coefficient at the 5% confidence level.
*** Significant coefficient at the 1% confidence level.

was expected for reasons mentioned above. On the other hand, this reform had a significant impact on the proportion of individuals staying at school at least until they turn 14 years old: 8% of the generation 1923 was forced to go on studying with the implementation of the Zay reform.

In a second step, we plugged the fitted value $\hat{E}(x_i|z_i)$ from the first stage estimation into the health equation. The return of one year of extra schooling β is estimated using the following probit model:

$$\begin{cases} y_i^* = \alpha + \beta\hat{E}(x_i|z_i) + \gamma z_i + v_i \\ y_i = \mathbb{1}_{y_i^* \geq 0} \end{cases}$$

Results of such an estimation are shown in Table 4.6. Like Wald estimators, school leaving age has a positive effect on survival at 50 years old. Nevertheless this coefficient is not significant at the 10% confidence level. Concerning the Zay reform, neither school leaving age nor the proportion of each generation still studying after

Table 4.6: Second stage estimations (probit)

	Zay reform y = survival at 80		Berthoin reform y = survival at 50
	Binary treatment	Variable intensity treatment	Variable intensity treatment
α	-0.094	3.43	0.075
	(0.13)	(4.41)	(1.51)
β	-0.32	-0.24	0.13
	(0.36)	(0.32)	(0.11)
γ	0.020	0.017	-0.014
	(0.016)	(0.015)	(0.012)

Standard errors in parenthesis. We subtracted 1900 to the year of birth. The Berthoin (resp. Zay) sample contains 47 337 (resp. 35 828) individuals born between 1950 and 1955 (resp. 1920 and 1925).

14 have a significant impact on survival at 80 years old. Eventually we do not find any causal impact of education on survival rate[5].

4.5 Discussion

Our empirical work is based on large samples containing around 40 000 individuals. Zay reform allows us to measure health at 80, which should be old enough to detect differences in mortality due to education if education did help to preserve one's health capital. Zay reform certainly did not increase much mean school leaving age, but it did force a significant share of the population to go on studying until 14. Berthoin reform indeed caused a sharp increase in school leaving age. Moreover 50 years old does not seem to be too young an age to reveal significant return to education on survival, since Lleras-Muney (2005) claimed to find a causal effect on mortality at similar ages. However, despite the large size of our dataset, we do not

[5]Other specifications have also been tested: separate regressions for men and women, and survival rates measured at different ages between 65 and 80 years old for the Zay reform. Results of these estimations, available upon request, are qualitatively similar to those presented in this paper. In particular, they do not show any significant causal impact of education on health.

find any causal impact of education on health.

These results are not in the mainstream of the existing literature, since a majority of studies showed a significant return to education on health. Among authors who specifically exploited temporal variations in mandatory education as a source of identification, four of them found a significant impact of education on health (Spasojevic 2003, Glied and Lleras-Muney 2003, Lleras-Muney 2005, Oreopoulos 2006) while Doyle et al. (2005) and Arendt (2005) did not. Adams (2002) had mixed results depending on the various health measures he considered. All these authors set up an instrumental variable framework. Moreover, Lleras-Muney (2005) implemented both an IV approach and a nonparametric regression discontinuity design. She showed a significant return to education only with the former framework, and explained away the non-significant RD results with imprecision in her computed death rates and a small sample size. We believe that RD designs are more relevant than IV when the exogenous variation in education is supposed to affect only individuals born after a given year. The reason is that changes in school leaving age, taken as an instrument, gradually lose some of their validity when the sample is extended to more birth cohorts. In particular the exclusion condition is no longer fulfilled if the sample is too wide, because year of birth may have a direct impact on health status, even when health is measured at the same age for everyone. Indeed health state of the overall population keeps bettering over time thanks to more efficient medicine and hygiene: those turning 50 today are in a better shape than 50 years old were a few decades ago. These improvements in health cannot be ignored when cohorts born ten or twenty years apart are in the same sample. As it is difficult to measure them, their effect on health is usually captured by the year of birth. This leads to a violation of the exclusion condition, since the simple fact of being born before or after the reform took place has a direct impact on health, besides the possible indirect one passing through education. Therefore education laws are a fully valid instrument only when used on a very small number of consecutive birth cohorts. An alternative approach could be to try to control for this generation effect by adding the appropriate function of year of birth in the health and education equations, which is more or less the essence of the parametric estimation we provided. The obvious difficulty is here to choose the relevant functional form. Few authors considered this issue, mainly because they often simultaneously faced the fact that health was not measured at the same age for everyone in their sample. As the structure of their dataset did not allow them to account for both effects, they chose to ignore the generation effect and focused on controlling the age effect by adding age as a covariate (Berger and Leigh 1989, Arkes 2003). It is likely that differences in health due to age are larger than those created by the generation effect; therefore their approach was certainly the correct one given their dataset. However, it would be more satisfactory to control for both sources of bias. Besides, assuming a linear dependence for the age or generation effects might be too restrictive, and a more flexible specification would be preferable. Lleras-Muney (2005) was able to include birth cohort dummies to account for the generation effect, but her model was identified solely because she took advantage of a very specific feature in her dataset covering all American states: changes in education laws did not happen the same year in all American states. Using the same source of identification on American data, Glied and Lleras-Muney (2003) opted for a flexible specification of the generation effect (birth cohort dummies) and a quadratic function of age. In general, changes in compulsory education were implemented at the country scale, which is the larger geographic area covered by usual datasets. In such a case it is impossible to separate the before/after reform dummy from the birth cohort dummies in an IV framework. Oreopoulos (2006) circumvented this issue on UK data by using quartic polynomial controls for birth cohort and age, instead of dummy variables. Our parametric estimation is similar to his. However, identification required that he specified a polynomial function for both birth cohort and age effects. We can avoid one of these two potential sources of misspecification, since our dataset enables us to measure health at the same age for everyone. By restricting our sample to only 6 birth cohorts, we limit the impact that (any misspecification of) the generation effect might have on the estimates.

Our results lead us to believe that education between 13 and 16 did not have any influence on subsequent mortality for those affected by Zay and Berthoin reforms. It is important to note that this does not rule out the existence of returns to education between 13 and 16 on other dimensions of health. Moreover, it may also be possible that years of schooling during early childhood or beyond the age of 16 might have a causal impact on mortality. Formal education may improve human capital in two dimensions which could in turn affect health status: knowledge and skills. Specific knowledge acquired at school may be useful for health production; this is certainly true for students who major in medicine or biology. But it is unlikely that knowledge acquired in classrooms between 13 and 16 may directly increase the ability to preserve one's health capital. According to official curricula in force in those days, only two courses in that range of age could bring specific knowledge relevant to health production: physical education and biology. Participating in physical activities at school may indeed influence attitudes and behaviors in everyday life: students could become more aware that physical activities may reduce health risks like heart disease or obesity. Several biology courses focussed on understanding how the human body works, with topics such as metabolic processes and internal regulation. The lesson on digestion emphasized the importance of choosing the right balance between nutrients, which may influence personal choices regarding nutrition. Lessons on cardiovascular and respiratory systems could provide scientific grounds towards more exercise. Besides the specific knowledge it provides, education may foster the development of cognitive skills, hence enabling better selection and integration of information from various sources. This dimension of human capital is certainly more relevant that knowledge when it comes to health production. General skills in reading, counting and understanding abstract ideas are indeed required for a proper comprehension of health related issues. Studying mathematics and science could create cognitive frameworks and reasoning tools useful for critical thinking. This last mechanism illustrates how education may lead to a more efficient allocation of inputs to produce health (Grossman 1972).

The impact of formal education on health is a cumulative process, which goes on through the grades. Each additional year at school should indeed contribute to improving health provided that it increased the amount of knowledge and skills that matter for health production. However returns to education might depend on student's motivation during schooling. Individuals obliged to stay at school may benefit less from an extra year of schooling than those voluntarily going on studying, both in terms of acquired knowledge and general skills. A key feature of Zay and Berthoin reforms is that they forced some people to study longer than they wished. This would imply that compliers indeed stayed longer at school but that this extra attainment did not significantly contribute to their intellectual growth. In particular, if the overall effect of education on health simply boiled down to the development of those skills useful for a deeper understanding of the world, returns to education on health as measured on compliers would then be smaller that the average return for always-takers (those who go on studying until 16 even in the absence of the Berthoin reform). In other words, both reforms might be weak instruments because they are not sufficiently correlated with dimensions of education that matter for health. The specific aspects of human capital relevant to health production may indeed react to some kinds of changes in education level, but not to changes in school leaving age. As education improves one's relative position in society, Ross and Mirowsky (1999) claimed that the more educated have a higher self-esteem, and hence a better health. In such a case, relative level of education compared to those who left school during the same year would be what really matters in terms of health. According to Grenet (2003), earnings in French labor market depended more on ranking in education level hierarchy than on the actual school leaving age in those days. As Zay and Berthoin reforms most likely did not alter this ranking in education, job opportunities of a given decile of school leaving ages distribution did not change even if their mean school leaving age increased. Hence the two reforms did not trigger the indirect causal mechanism linking education and health through an income effect. In the same vein, possible causality passing through working conditions (safer or less wearing jobs for the more educated) cannot be tested with our instruments.

The absence of a significant causal impact in our study may then be due to the choice of school leaving age as a proxy for education level. As stated above, school leaving age may be a poor proxy for knowledge, skills and social position in society, all of which may have an influence on health. This could explain why we do not find any causal impact of education on health. To put these possible limitations into perspective, one must have in mind that Lleras-Muney (2005) claimed to find a causal impact of education on health in a very similar set up: she used school leaving age as education proxy and mortality as health measure, her instruments were changes in compulsory education, and her compliers were students leaving school around 14 years old.

4.6 Conclusion

Correlation between education and health is well documented in epidemiological literature. There is a strong empirical link, and this holds with various measures of health: self-rated health, morbidity, mortality, and physical limitations. However, possible mechanisms creating that correlation are still largely unknown. In the last years a growing number of studies have focused on determining whether causal mechanisms could explain part of that correlation. As education level is an individual choice depending on unobserved characteristics, an endogeneity problem arises when these characteristics also affect health status. Such variables could be familial background or time preferences. A common strategy to deal with this issue is to exploit exogenous variations in education. Such variations in education are not correlated with other variables influencing health, and thus it is possible to test the existence of a causality running from education to health. Changes in compulsory education laws are a good candidate for such exogenous shocks, and they are widely used to estimate both monetary and non-monetary returns to education. We exploit two changes in mandatory schooling in France, which increased minimal school leaving age by one year in 1923 and two more years in 1953.

Authors using that kind of reform usually estimate return to education in an instrumental variable framework. The first contribution of our paper is to use a regression discontinuity design, which is more appropriate to the nature of such legal changes. Moreover assumptions required for identification are weaker than in the instrumental variable approach. In particular it allows us to provide a nonparametric estimate of return to education. Our second main contribution is that health status is measured at the same age for everyone. As most empirical papers work on cross-sectional datasets, individuals in their sample do not have the same age when their health is measured. Therefore they must take into account a possible age effect on health if they want to estimate an unbiased return to education. Working on survival rates at a constant age is a convenient way to be sure that differences in health are not due to any misspecification in the health dependence in age.

Our results indicate that survival rates at 50 and 80 years old do not seem to be affected by years of schooling between 13 and 16. Besides the possibility that education might not have a causal impact on health, several reasons could account for this absence of significant return to education in our study. The most credible explanation is certainly the choice of school leaving age to measure one's level of education. It is indeed not clear that changes in school leaving age would capture the changes in human capital which are relevant to health production. However, most papers in the literature face the same possible limitations, and some of them claim to find significant returns to education on health nonetheless.

Chapter 5

Returns to education during World War II

Sommaire

5.1 Introduction

Education and health are considered two of the main determinants of human capital, and for this reason they were first used as explanatory variables in wage equations. Noticing a strong correlation between those two variables, Grossman (1972) then built a theoretical framework modeling their interdependence. He postulated the existence of a health production function; education was one of its inputs and hence could affect health status. As such, Grossman was describing a causal link from education to health. However, this mechanism is only one of the three possible explanations to the empirical correlation we observe between education and health. The correlation may also come from a causality running in the opposite direction, with health having a causal impact on education (Kremer and Miguel 2004). Eventually, this correlation may not be based on any causal relationship between education and health. It would stem from a third factor influencing both education and health, and thus indirectly creating a link between them. In such a case, a plausible candidate would be one's familial background: parents who stimulate their children to achieve successful studies may at the same time insist on the importance of behaving healthily. Time preferences may also be an unobserved characteristic simultaneously governing education and health choices, since investing in education and in health improves one's wellbeing mostly in the long run.

A growing literature focuses on determining whether any causal relationship from education to health could explain, at least partly, the empirical correlation. The most common strategy to deal with potential endogeneity of education is to use an instrumental variable (IV). The idea behind IV is to isolate a specific event which induced a variation in educational level. As long as this event is not correlated with variables suspected of influencing both education and health, its sole possible way to affect health status is to trigger an indirect mechanism passing through education. Therefore this method can reveal causal paths running from education to health, when such paths do exist. Several types of instruments have been used in such a case. The most popular is certainly the exogenous variation in education created by a change in compulsory education laws (e.g. Arendt 2005, Oreopoulos 2006, 2007, Silles 2009, van Kippersluis et al. 2009). Adams (2002) uses quarter of birth, which is correlated with the number of years spent in school, and should not have any direct impact on health. Arkes (2003), Auld and Sidhu (2005) exploit local unemployment rate as an instrument. De Walque (2007) uses the Vietnam draft lottery, which induced some students to extent their schooling to avoid the draft.

Though a majority of studies claim to find a causal impact of education on health, no consensus has emerged yet. This lack of consistency could stem from the variety of health measures used in the literature, if education was supposed to affect only specific dimensions of health. This hypothesis does not seem to be valid, since literature provides contradictory results on causality with a given health measure (e.g. Arkes (2003) and Oreopoulos (2006) on activity limitations, or Arendt (2005) and Adams (2002) on self-rated health). Mortality is no exception to the rule. Cipollone et al. (2006) rely on exemption from military service after an earthquake for identification and find no causal impact on mortality. Using compulsory schooling laws, Lleras-Muney (2005) and Glied and Lleras-Muney (2008) on US data and van Kippersluis et al. (2009) on Dutch data report a significant causal effect of education on mortality. On the other hand, Clark and Royer (2008) on UK data and Albouy and Lequien (2009) on French data claim that raising minimum school leaving age did not induce any significant decrease in subsequent mortality for those affected by those reforms[1].

This study is the first to provide evidence of a significant causal link between education and mortality on French data. We propose a two-stage estimation using an innovative instrumental variable: France was directly involved in World War II, and was invaded by the German army during the conflict. War and several years of foreign occupation certainly had a sizeable impact on everyone's health. However, it also led to a decrease in school leaving age for a specific subpopulation, namely those who left school during the war. As long as the direct impact on health was the same for everyone in our sample[2], this exogenous drop in education allows us

[1] See Grossman (2004) or Cutler and Lleras-Muney (2008) for a more comprehensive survey, as well as a discussion on the various measures of health used in the existing literature.

[2] Section 5.5.2 discusses the validity of this assumption.

to identify a causality running from education to health. We find that one extra year of schooling increases the probability of being still alive at 60 by 6 percentage points, among those affected by our instrumental variable.

The next paragraph describes the historical context, and section 5.3 presents the dataset as well as basic statistics. Section 5.4 shows our results, followed by a discussion on the validity of the instrument. We conclude with a possible heterogeneity of returns to education due to students' motivation at school. Comparing Albouy and Lequien's (2009) results with ours indeed suggests that returns are positive and significant for motivated students, but not significant for students forced to comply to mandatory education laws.

5.2 Historical background

World War II had a dramatic impact on countries involved in the conflict. Millions of people died, both military forces and civilian society were burdened with these losses. Numerous European cities were destroyed by aerial attacks, schools were also frequently targeted. Everyday life during the five years of war was severely disrupted, and it is likely that students in these days faced more difficulties to reach their optimal level of education. Ichino and Winter-Ebmer (2004) were the first to rely on this idea, when they estimated the educational cost of World War II in terms of wages. Their strategy was based on the comparison of wages between two countries which were directly affected by the war (Germany and Austria) and two non war countries (Switzerland and Sweden). Individuals in their sample were born between 1920 and 1950, and wages were measured in the 1980ies. Though properly set up, their work suffers from a drawback common to all cross-country studies: they have to assume that all possible events affecting wages between the World War II and the 1980ies had a similar effect on wages in these four countries.

We are able to relax this strong assumption by working on a single country, France. Moreover, we take advantage of a very specific feature of the conflict in that country which probably magnified the impact of war on French society: on top of the usual disruptions observed in other countries at war, France lived under foreign military occupation during most of the war.

War and occupation affected every aspect of French people lives[3]. There are several reasons to believe that war lowered education by disorganizing school environment: schools may have closed, teachers' attendance may have declined, parents may have fled the conflict and not necessarily sent their children to school in their new town. Besides, war probably had an impact on time preferences, because immediate matters, such as managing to stay alive until the end of the day, took precedence more strongly than before over long-term projects like education. One also had to go through hardships like fear of being arrested, and more generally all kinds of worries brought by a foreign military invasion. Since relative importance of short-term and long-term investments was altered in decisions making processes, students in those days were likely to leave school earlier than they would have if there had been no war. We use the disruptions caused by World War II and German occupation as an exogenous source of variation in school leaving age.

5.3 Data

5.3.1 Descriptive statistics

The empirical analysis is based on the Permanent Demographic Sample (Echantillon Démographique Permanent, EDP thereafter), a longitudinal dataset containing information on all women and men living in France and born on one of the first four days of October. It compiles 1968, 1975, 1982, 1990 and 1999 census data with information from register of births, marriages and deaths from 1968 until 2005. Our estimation sample is an extract from the EDP, from which we selected all individuals born between 1920 and 1945. As the EPD is currently updated until 2005, 60 is the oldest age at which we know for everyone in our sample whether one is alive or not. Then we used the date of death to exclude those who died before turning 48 years old[4]. Table 5.1 presents descriptive statistics of this sample. For

[3]See Sirinelli et al. (2004) for a detailed description of living conditions in France during the war.

[4]This sample selection, as well as the choice of this health measure, were dictated by specific features of the dataset. Section 5.5.1 details the reasons leading to those choices and provides a

Table 5.1: Descriptive statistics

	Total sample	Women	Men
Number of observations	147 780	75 442	72 338
Proportion of women	0.511		
	(0.50)		
Year of birth	1932.13	1932.08	1932.19
	(7.46)	(7.46)	(7.46)
School leaving age	15.61	15.45	15.78
	(3.01)	(2.73)	(3.27)
Survival at 60	0.92	0.95	0.89
	(0.27)	(0.20)	(0.32)

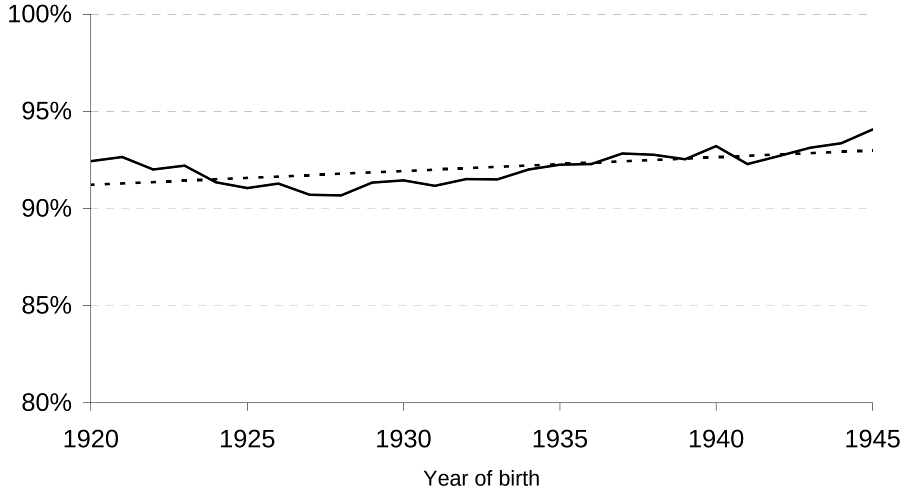
Notes: Standard errors in parenthesis. Total sample consists of individuals who were born between 1920 and 1945.

those 147 780 individuals who survived up to 48, we measure health status with a dummy indicating whether one is still alive at 60. Death certificates come from an exhaustive dataset at the country level, hence we have information on the exact date of all deaths occurring in France. Unlike most studies on mortality, we observe the date of death at the individual level, and so we don't have to approximate survival by survival rates at the cohort level (see van Kippersluis et al. 2009). Moreover the longitudinal structure of our dataset allows us to compute a health measure at the same given age from everyone in our sample. Therefore, we do not face the issue of specifying how age influences health status.

According to Elo and Preston (1996), differences in mortality due to socioeconomic status appear before 55, and begin to fade away after 65. Therefore 60 is a relevant age to show causal effects from education to health, if these effects transit through changes in socioeconomic status. Moreover, as an objective measure of health status, survival rate does not suffer from bias inherent to all self-declared measures. Figure 5.1 plots survival rate at 60 years old by birth cohort. There is a slightly increasing temporal trend, accounting for gradual improvements in health over time (due to medicine progress, better living conditions, etc.). Moreover, sur-

discussion on their consequences.

Figure 5.1: Survival rate at 60 years old by birth cohort



Note: Proportion of each birth cohort still alive at 60 years old, conditionally on being alive at 48 (plain line). The dotted line is the corresponding linear fit. Based on the 147 780 individuals who were born between 1920 and 1945.

vival rate is lower for birth cohorts 1924 to 1934.

For each individual in our sample, censuses provide us with her/his school leaving age *sla*. Figure 5.2 represents the distribution of this school leaving age. More than 40% of the population left school at exactly 14 years old. This spike corresponds to pupils reaching the age at which school was no longer mandatory[5]. 14% declare that they left school before reaching 14, which was theoretically not allowed. This is probably due to misreporting, or imperfect compliance with the law. A similar phenomenon is observed in other countries (Silles 2009).

Figure 5.3 pictures the mean school leaving age by generation and its linear

[5]Compulsory education laws were modified twice in France during the first half of the twentieth century. Education used to be mandatory from 6 until 13 years old, for children born before 1922. It was then extended in 1933 to 14, this reform being effective for children born after January 1^{st}, 1923. However this reform did not have a significant impact on average school leaving age, since most of the population was already going to school up to 14 when it came into force. Another reform then set to 16 the minimum school leaving age, for those born after 1953. See Albouy and Lequien (2009) for more details.

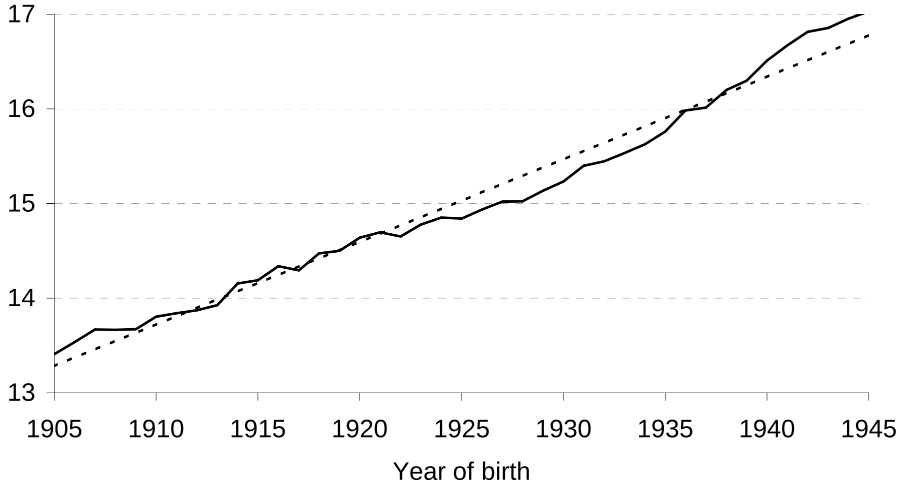
Figure 5.2: Distribution of school leaving ages

Note: Proportion of the sample leaving school at a given age. Based on the 147 780 individuals who were born between 1920 and 1945.

fit. We notice the well-known positive trend over time, stating the rise in level of education through generations. There is a drop in education for cohorts 1922 to 1934. The 1934 birth cohort was only 11 years old at the end of the war in 1945, and yet seems to have been significantly affected by the war. One must have in mind that the magnitude of the drops does not represent the actual effect of war on one's level of education, since only a fraction of each birth cohort left school during the war. Cohorts were unevenly affected, the intensity of the impact being related to the proportion of each generation still studying during the war. Hence decreases in Figure 5.3 can only give a hint of the differential impact between cohorts.

The empirical correlation between education and health is shown in Figure 5.4. A positive link is clearly visible: the longer one stays at school, the likelier one is to be still alive at 60. The aim of this study is to disentangle a possible causality from this correlation. Our identifying assumption is related to the 10-year long drops in survival rate (Figure 5.1) and education (Figure 5.3) happening for roughly the same birth cohorts. If the war is indeed a valid instrument, then the decrease in survival rate is a consequence of the drop in education: war had an indirect effect

Figure 5.3: Mean school leaving age by birth cohort



Note: Mean school leaving age as a function of the year one was born (plain line). The dotted line is the corresponding linear fit. Based on the 212 614 individuals who were born between 1905 and 1945.

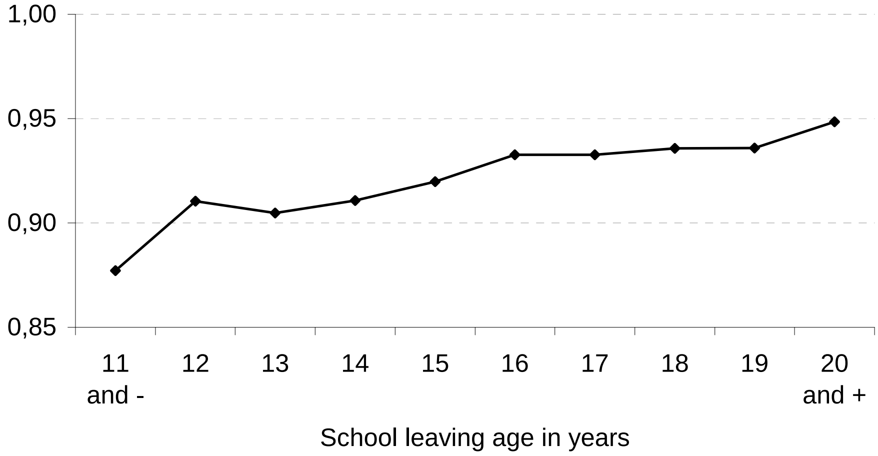
on health through its impact on education.

5.3.2 War as an instrumental variable

Following Ichino and Winter-Ebmer (2004), the instrumental variable is a dummy equal to 1 for those belonging to birth cohorts whose level of education was the most affected by the war. These birth cohorts are those with the higher share of individuals giving up school between 1939 and 1945. It is indeed likely that the war did not have any impact on the decision to leave school for those who dropped out before the war broke out, since these children could not foresee the war when they decided to leave school. In particular, students were not induced to leave school before the war because they anticipated that the educational system would suffer from the war.

As 40% of each birth cohort left school at 14 (Figure 5.2), we define the range of birth cohorts affected by the war by selecting those who turned 14 during the war. Hence the instrumental variable $war1$ is a dummy set to one for those who were

Figure 5.4: Correlation between school leaving age and probability of survival at 60



Note: Survival at 60 as a function of the school leaving age. Based on the 147 780 individuals born between 1920 and 1945.

born between 1925 and 1931, to 0 otherwise. The impact of the war on education can vary with gender through an interaction with sex.

The next section presents the econometric framework, along with estimation results using $war1$ as instrumental variable. As robustness checks, section 5.5.3 provides alternative definitions of the instrument and results of the corresponding estimations.

5.4 Results

We aim at estimating the impact of education on health. If there were no unobserved heterogeneity altering both education choices and health, estimation of probit model (5.1) would give us an answer:

$$\begin{cases} surv_i^* = a_1 + b_1\ sla_i + c_1\ sex_i + d_1\ year_i + \alpha_i + u_{1i} \\ surv_i = \mathbb{1}(surv_i^* \geq 0) \end{cases} \tag{5.1}$$

Health is measured by $surv_i$, a dummy variable equal to 1 if individual i is still alive at 60. Its corresponding latent variable is $surv_i^*$. First explanatory variables are school leaving age sla_i and gender ($sex_i = 0$ if i is a woman, 1 otherwise). α_i is a geographical fixed effect, based on the region of birth. It should capture structural differences in health existing before the war between the 22 administrative regions in France, as well as the effect of war on health common to all inhabitants of a given region. We use the year of birth $year_i$ to take into account any progressive improvement in medicine or public health which would affect survival. u_{1i} sums up all other (un)observed variables.

If education choices were not endogenous, this error term would not be correlated with explanatory variables, and a probit estimation would be convergent. Results of such an estimation are presented in column 1 of Table 5.2. Men have a lower survival rate than woman, as Table 5.1 indicates. Moreover, one extra year of education is associated with an increase by 0.5 percentage point of the probability of being still alive at 60. Results are qualitatively the same when we allow the impact of region on health to be different for men and women through an interaction between region of birth and sex_i (column 2 of Table 5.2).

In order to account for the possible endogeneity of education, the first stage of our instrumental variable strategy consists in introducing an exogenous source of variation among variables explaining education:

$$\begin{aligned} sla_i = {} & a_2 + b_2\, sex_i + c_2\, war1_i + d_2\, war1_i \times sex_i \\ & + e_2\, year_i + \beta_i + u_{2i} \end{aligned} \tag{5.2}$$

Level of education sla_i depends on gender, whether one belongs to a birth cohort affected by the war $war1_i$, and an interaction term between $war1_i$ and sex_i. β_i is a regional fixed effect, which can be interacted with sex_i in some specifications. A temporal trend is also present, captured by $year_i$. As can be seen in columns 1 and 2 of Table 5.3, men stay longer than women at school. This was expected, since men have a higher mean school leaving age (Table 5.1). It appears that the instrumental variable $war1$ (being born between 1925 and 1931) has a significant impact on education: individuals belonging to cohorts affected by the war drop out sooner, and the difference is more pronounced for men (-0.19 year) than for woman

Table 5.2: Probit estimations

	Dependant variable	
	Survival at 60	Survival at 60
	(1)	(2)
sla	0.0052***	0.0053***
	(0.00025)	(0.00025)
sex	-0.070***	-0.071***
	(0.0014)	(0.011)
year	0.00012	0.00011
	(0.00009)	(0.00009)
region	yes	yes
$region \times sex$	no	yes
N	147 780	147 780

Notes: Estimations are run on individuals who were born between 1920 and 1945. We subtracted 1910 from the year of birth *year*. Regional fixed effects are not reported. Columns (1) and (2) are probit regressions, and coefficients are mean marginal effects. Robust standard errors in parenthesis. *, **, and *** mean that coefficients are significant at respectively 10%, 5% and 1% confidence levels.

Table 5.3: IV Probit estimations

	First step School leaving age		*Second step* Survival at 60	
	(1)	(2)	(3)	(4)
intercept	13.27***	13.34***		
	(0.061)	(0.082)		
$\widehat{sla}$			0.066***	0.066***
			(0.013)	(0.0013)
sex	0.33***	0.19*	-0.10***	-0.095***
	(0.017)	(0.11)	(0.0045)	(0.015)
*war*1	-0.14***	-0.14***		
	(0.022)	(0.022)		
*war*1 × *sex*	-0.052*	-0.057*		
	(0.029)	(0.029)		
year	0.099***	0.099***	-0.0060***	-0.0060***
	(0.0011)	(0.0011)	(0.0013)	(0.0013)
region	yes	yes	yes	yes
region × *sex*	no	yes	no	yes
N	147 780	147 780	147 780	147 780

Notes: Estimations are run on individuals who were born between 1920 and 1945. We subtracted 1910 from the year of birth *year*. Regional fixed effects are not reported. Columns (3) and (4) are probit regressions, and coefficients are mean marginal effects. Robust standard errors in parenthesis. *, **, and *** mean that coefficients are significant at respectively 10%, 5% and 1% confidence levels.

(-0.14 year).

We then use the predicted school leaving age $\widehat{sla_i}$ from (5.2) into the second step of our estimation:

$$\begin{cases} surv_i^* = a_3 + b_3\,\widehat{sla_i} + c_3\,sex_i + d_3\,year_i + \alpha_i + u_{3i} \\ surv_i = \mathbb{1}(surv_i^* \geq 0) \end{cases} \tag{5.3}$$

Results are shown in columns 3 and 4 of Table 5.3. The coefficient on gender is in line with aggregate statistics presented in Table 5.1: survival at 60 is lower for men than for women (-0.1). This is a well-known fact about the French population, observed even prior to the war. The magnitude of the temporal trend is low, and is quite surprisingly negative. A possible explanation could be that war had a direct age-specific impact on health, and that young children during the war were more adversely affected than adults by starvation and stress. Section 5.5.2 discusses this issue and shows that it does not necessarily invalidate the exclusion condition. Education has a positive and significant causal impact on survival: one additional year of education increases the probability of being alive at 60 by 6 percentage points. This coefficient is highly significant and means that education has a positive and significant causal impact on compliers' health.

The IV-probit coefficient is larger than the one obtained by probit, and those two estimates are statistically different (a Wald test of exogeneity gives $\chi^2(1) = 37.46$, $pvalue < 0.0001$). This confirms that education is indeed an endogenous variable in the health equation. On the other hand, one would have expected the IV coefficient to be smaller than the probit one. Silles (2009) also finds that IV estimates are several times larger than OLS estimates on UK data. This may be due to a direct effect of our instrument on the outcome *surv* (see section 5.5.2). Another possible explanation is that education is measured with error, which biases downwards the probit estimate. This bias disappears in IV estimations if the instrument is independent of the measurement error, which is likely to be the case here[6]. Eventually, the difference between probit and IV-probit estimates may be due to the instrument, which identifies returns to education on a very specific subpopulation with high returns (see section 5.6). Ichino and Winter-Ebmer (2004) indeed find high monetary returns in Austria and Germany with a similar instrument, and therefore a similar set of compliers.

[6]See Grossman (2004) or Card (2001). See also Belzil (2007) who provides an explanation using a dynamic framework.

5.5 Discussion

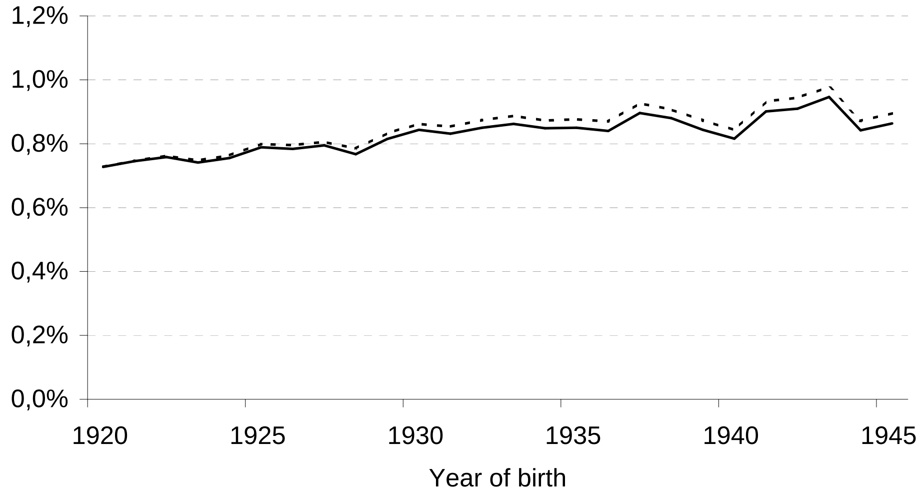
5.5.1 Sample selection

The EDP was created from the 1968 national census, which means that individuals who died before 1968 are not present in this dataset. It is therefore possible that the EDP is not a perfectly faithful picture of the French population living during World War II.

A possible cause of selective mortality before 1968 could be that older individuals in our sample were more likely to become soldiers during the war and to die on the battlefield. Severe health problems during childhood caused by the war could also lead to non random sample selection. Kannisto et al. (1997) and Roseboom et al. (2001) claim that mortality rates at 17 and 18 years old are higher for those who suffered from acute starvation during early childhood, compared to children older during the famine (see also section 5.5.2). This would imply that the possible sample selection may not have the same magnitude in each birth cohort. On the other hand, they also claim that mortality at higher ages does not depend on the age one suffered from starvation, once mortality early in life is taken into account. Following them, we tried to keep constant any possible selection bias throughout the sample, and excluded individuals who died before reaching adulthood. As everyone in our estimation sample was born after 1920 and we have information on death from 1968 on, the youngest possible age for such a truncation is 48. Therefore, our health measure, being alive or not at 60 years old, must be understood as a survival index given that one is still alive at 48.

It is possible to quantify the extent of selective mortality prior to 1968, using the yearly number of births in France from the register of births, marriages, and deaths (Insee 2007). Figure 5.5 represents the proportion of individuals born one given year in France who are present in our estimation sample. Separate curves are plotted depending on whether those who died before turning 48 were excluded or not from the numerator. The EDP contains individuals born on one of the first four days of October. So the proportion of individuals present in the EDP should theoretically be close to 4/365 ($\approx 1.1\%$) if everyone had survived up to the creation

Figure 5.5: Sample selection



Note: For each year of birth, number of individuals in our estimation sample divided by the total number of births in France that year (plain line). The dotted line represents a similar ratio, where we did not exclude those who died before turning 48 from our sample.
Sources: EDP and Insee (2007).

of the EPD in 1968. Figure 5.5 indicates that our estimation sample contains about 0.8% of the selected births cohorts. A slightly increasing trend is visible, certainly accounting for gradual improvements in public health and medicine over the years. Figure 5.5 shows no clear sign of selective mortality in our sample due to more frequent enlistments in older cohorts. Hence the composition of our sample does not seem to have been affected by possibly different proportions of soldiers across cohorts.

There is a small gap from 1938 to 1940 on both curves, which means that a relatively smaller share of these births cohorts was still alive in 1968 (dotted line) or at 48 years old (plain line). So selective mortality in early life seems to be more important for the youngest cohorts during the war, which is in line with what Chen and Zhou (2007) observed after the great famine in China (see section 5.5.2). The magnitude of this gap is low, around 0.05%, and therefore this possibly non random selection should not jeopardise our estimations.

Another specificity of the EPD is the availability of death certificates. Although they are exhaustively collected for deaths happening in France, most death certificates are missing for individuals who died while abroad. It means that these individuals are wrongly considered to be alive in our estimations. However the number of such deaths is certainly small compared to all deaths occuring in France. Besides, this measurement error would bias our estimations only if dying abroad after 1968 were correlated with our explanatory variables (including the instrument), which seems unlikely.

5.5.2 Validity of the instrument

Like in any other study, an instrumental variable must satisfy two criteria to be valid. The first one is to be correlated with the endogenous variable, which here is the school leaving age. This is indeed the case, as shown in Figure 5.4. The second condition is usually trickier, because its validity can only be theoretically assessed. It stipulates that the instrument must not be directly linked to the variable of interest, health. In other words, the simple fact of belonging to cohorts whose education choices were affected by the war must have no other impact on subsequent health that the indirect one due to a lower education. It is obvious that everybody in our sample suffered from the war and its consequences (lack of food, stress, etc.). This certainly had a sizeable impact on their health. The exclusion condition requires that the effect of all possible events altering health, except war through its impact on education, did not depend on the age individuals had during the war.

5.5.2.1 Compulsory labor service

Few mechanisms could have affected health in an age-varying way. The first to cross one's mind is the compulsory labor service (Service du Travail Obligatoire, STO). In 1942 Germany lacked labor supply and asked the French government to send 250 000 men to Germany. It was first set up on a voluntary basis, but failed to recruit the required number. Therefore it became mandatory in 1943, the enrollment mainly concerning men born between 1920 and 1922 (Sirinelli et al. 2004). As we ran our estimations on people born between 1920 and 1945, our sample contains the cohorts that were the most affected by STO. Restricting the sample to cohorts 1923-1945 does not qualitatively change our estimation results. However, even with this restricted sample, it is still possible that some individuals among the oldest in our sample were indeed forced to go to Germany[7].

Except maybe for the minority sent to farms in the German countryside, living conditions (including food supply) faced by STO workers were certainly worse than in France. So their average health was not likely to improve during their stay abroad, compared to those who remained in France. STO should then reduce survival differences, if they ever existed, indirectly caused by the instrumental variable: those whose survival may lower because war forced them to drop out of school were too young to be sent to Germany, whereas some people among those who left school before the war enrolled in STO and so have lower survival rates as well[8]. Hence this direct impact of STO on health might bias downwards any possible causal effect of education on survival. But this will not invalidate the causal interpretation we suggest, should we find a positive and significant impact.

Besides this possible direct effect on health, STO may also have indirectly lowered education during the war, if fourteen-year-olds dropped out of school to replace in plants older workers sent to Germany. As such, STO could be described as one specific channel through which war could have decreased mean school leaving age. This channel is likely to satisfy the exclusion condition.

[7]They had most likely left school before the war (more than 80% of birth cohort 1923 had left school before 1939). As for the few still studying, we are not aware of any law stipulating that students were not eligible to STO. So there was no reason for them to go on with their studies in order to avoid STO. This rules out strategic behaviors similar to those of the Vietnam draft lottery, when American students attended school longer to avoid being sent to the Vietnam war (see Angrist 1990, De Walque 2007). In particular, it implies that STO does not invalidate the monotonicity assumption formulated by Angrist and Imbens (1994). This assumption is required in order to give a causal interpretation of the estimated return to education as an average treatment effect on a specific subpopulation (see section 5.6).

[8]Here we implicitly assume that education has a non negative impact on health.

5.5.2.2 Enlistment

Another possibility is that students induced to leave school early became soldiers. They could either enlist in what remained of the regular army (some troops were stationed in North Africa), or join the resistance. Besides possible selective mortality during the war (see section 5.5.1), several studies suggest that becoming a soldier may have an impact on later health. Bedard and Deschenes (2006) claim that military service caused an increase in smoking prevalence among World War II and Korean war veterans, which in turn had a long-term effect on their health. Being a war veteran might also negatively affect later wages (Angrist 1990 for Vietnam veterans, Angrist and Krueger 1994 for World War II veterans). By granting a better access to health care, income may have an impact on health; hence becoming a soldier could also affect health through changes in wages.

In our context, the exclusion condition does allow veteran status to have a direct impact on later mortality. But if this impact indeed exists, the proportion of veterans has to be the same in the set of birth cohorts affected by the instrumental variable $war1$ (1925-1931) and those not affected (1920-1924 and 1932-1945). Generations 1932-1945 were obviously too young during the war to become soldiers; there were certainly more soldiers in birth cohort 1920-1924 than in 1925-1931. However, this higher proportion in oldest cohorts may not totally offset the absence of veterans in the youngest cohorts. Therefore a higher share of veterans among those affected by the instrument seems plausible. Provided that being a soldier indeed decreases health, this would violate the exclusion condition. Estimates of returns to education would then be biased upward. We are not aware of any evidence suggesting that this difference in proportion might be large. It is obviously difficult to know precisely how many individuals joined the resistance. Since 1945, the Ministry of Veretans has officially recognized around 130 000 resisters. Adding former deportees and those who died during the war, about 2% of the French adult population is considered to have played an active role in the resistance (Wright 1962, Paxton 1982). So only a minority is supposed to have joined the resistance, and such a small proportion in the overall population is unlikely to result from large differences in proportion between birth cohorts. Therefore we believe that any violation of the exclusion condition caused by the resistance should not lead to substantial bias in our estimations.

5.5.2.3 Starvation

Food rationing could also have altered health with a varying intensity in our sample. France indeed faced a severe lack of food supply during the war, and it is likely that starvation had more or less serious consequences on health depending on the age at which individuals suffered from it.

There is mixed evidence that adverse events in early life would affect old age health (Costa 2003 versus Cutler et al. 2007)[9]. Among studies focusing on whether the magnitude of this possible impact would vary with age during adverse events, several work on long-term effects of famines. These are especially relevant to our study, since famines created extreme living conditions quite similar to those existing in France during the war. From that point of view, they constitute a better benchmark than studies realized during peacetime or without any major natural disaster. Chen and Zhou (2007) use the 1959-1961 great famine which occurred in China. They find that adult health, measured by attained height, was negatively affected by the famine for those who suffered from the famine during childhood. Moreover the youngest ones (birth cohorts less than 3 years old during the famine) were more strongly impacted than older children. Kannisto et al. (1997) exploit the 1866-1868 Finnish famine. They focus on individuals born between 1861 and 1873, and check whether survival rates at ages ranging from 17 to 80 depend on birth cohort. They observe that survival at 17 is lower for birth cohorts born during or right before the famine. However, survival rates at 40, 60 and 80 years old are not significantly different among birth cohorts, once mortality before 17 is controlled for. A similar methodology applied to the 1944-1945 Dutch famine indicates that survival at 50 (Roseboom et al. 2001) and 57 years old (Painter et al. 2005) do not seem to depend on prenatal or early childhood exposure to the famine, once mortality during the

[9]See Case et al. (2005) for the impact of *permanent* childhood characteristics, like family socioeconomic status, on various adult outcomes, including health.

18 first years of life is taken into account.

These results suggest that extreme living conditions during childhood would not have any direct age-varying effect on adult survival, given that one did not die during childhood. This is comforting, because our health measure is precisely a survival rate conditional on being still alive at 48. Nevertheless there is currently no definitive evidence in the literature, and we cannot rule out that the war might have had a direct age-specific effect on survival at 60 years old. Such a mechanism would violate the exclusion condition. Returns to education would then be overestimated if the war had indeed a stronger negative impact on health for birth cohorts affected by the instrumental variable, and vice versa. According to Chen and Zhou (2007), young children during the war were the most likely to suffer more intensely than others from the war, and the instrument did not affect these birth cohorts. As a consequence, any direct age-specific effect of war on health would lead to underestimate the impact of education on health. This strengthens our conclusion since we find positive and significant returns to education despite this possible bias.

Eventually we find no decisive explanation undermining the validity of our instrumental variable.

5.5.3 Robustness checks

Identification comes from exogenous changes in education for a part of our population (the compliers). Hence a crucial step is to define which individuals were induced to modify their education choices because of the war. Like in many studies based on a instrumental variable, two reasons forbid us to partition the population into two groups, depending on whether one was affected by the instrument. Being affected by the war is not a binary treatment, and the intensity of its impact varies continuously in the population. Moreover it is not possible to know the exact list of individuals affected by the instrument, since our dataset does not contain that information. Nevertheless identification is possible if we define two groups of individuals, the first one being on average more intensively affected by the instrument than the other. That amounts to selecting birth cohorts about to reach their school leaving age during the war (Ichino and Winter-Ebmer 2004). Results presented in

Table 5.4: Returns to education with alternative instruments

Instrument	With $region \times sex$	Without $region \times sex$
$war1$	0.066***	0.066***
	(0.013)	(0.013)
$war2$	0.052***	0.0542**
	(0.0086)	(0.0086)
$war3$	0.079***	0.079***
	(0.015)	(0.015)
$war4$	0.057***	0.057***
	(0.0095)	(0.0096)
$war1$, Zay	0.062***	0.062***
	(0.010)	(0.010)
$war2$, Zay	0.054***	0.054***
	(0.0084)	(0.0084)
$war3$, Zay	0.057***	0.057***
	(0.011)	(0.011)
$war4$, Zay	0.058***	0.058***
	(0.0094)	(0.0095)
N	147 780	147 780

Notes: Each coefficient corresponds to the return to education given by equation 5.3, where $war1$ has been replaced by another instrument. Estimations are run on individuals who were born between 1920 and 1945. We subtracted 1910 from the year of birth $year$. Column (1) includes an interaction between regional fixed effects and sex in the set of regressors, column (2) does not. Robust standard errors in parenthesis. *, **, and *** mean that coefficients are significant at respectively 10%, 5% and 1% confidence levels.

section 5.4 are based on the following criteria (see section 5.3.2): the most affected individuals are those who turned 14 during the war, because more than 40% of each birth cohort left school at 14 in those days. With this definition, the instrumental variable $war1$ is a dummy set to one for those who were born between 1925 and 1931, to 0 otherwise.

However, the decision to leave school in the aftermath of the war might have been affected by the war itself. Indeed education supply was not instantly back in 1945 to its pre-war level: it took some time to rebuild schools and find new teachers. Moreover labor demand certainly increased after the end of the war, because workers were needed to rebuilt destroyed plants and towns. This could have induced some students to reduce their optimal level of education after the war was over. Figure 5.3 seems to indicate that such behaviors indeed happened and that it concerned birth cohorts up to 1934. Taking into account that education choices might have still been disrupted in the aftermath of the war, a second definition adds 1932-1934 to the set of cohorts affected by the instrument ($war2$).

According to Figure 5.2, around 60% of each birth cohort left school between 14 and 16. So in order to include cohorts with a majority of individuals likely to be constrained in their education choices, another possibility is to choose cohorts 1923 to 1931 ($war3$). This corresponds to individuals who were between 14 and 16 years old during the war. Once again, an alternative definition adds cohorts 1932-1934, to also cover those between 14 and 16 in the three years after the war ($war4$).

As robustness checks, we estimated the econometric model described in section 5.4 with $war2$, $war3$, or $war4$ instead of $war1$. We also used interactions between these four instruments and sex as additional instruments. As minimum school leaving age was increased from 13 to 14 for those born after 1923 (see footnote 5), we added this exogenous shock on education in our set of instruments (named Zay after the Minister of Education at that time, Jean Zay). Finally, we ran all these regressions with and without the regional fixed effects interacted with sex. None of these changes qualitatively affects our results (Table 5.4): the magnitude of returns to education goes from 5% to 8% depending on the specification, and the coefficient is statistically significant at the 1% confidence level.

5.6 Heterogeneous returns to education: Voluntary versus mandatory schooling

Although we do find a significant causal effect of education on health, we do not know precisely how a higher level of education could lead to a better health. For instance, education could improve human capital, and these improvements could then help to cope with health related issues. Specific knowledge acquired during biology courses could indeed be useful to understand how the human body works.

Moreover, education could stimulate intellectual growth and increase one's ability to grasp abstract ideas. The more educated seem more likely to comply with complex medical treatments (Goldman and Smith 2002), to take advantage of available information (De Walque 2004, Chen and Lange 2008) or technological innovation in medicine (Glied and Lleras-Muney 2008). Education may also improve human capital in dimensions which do not directly matter for health (Oreopoulos and Salvanes 2009), but some are relevant to productivity in the labor market. The more educated would be more productive, and then earn higher wages. This could indirectly, but still causally, affect health status, since they would have a better access to costly treatments and preventive care. This multiplicity of possible causal paths could create differences in the magnitude of returns to education across individuals, depending on which causal mechanism is triggered.

Among other factors accounting for possible heterogeneous returns, the benefit from one extra year of schooling may depend on the age one leaves school (Aakvik et al. 2010, Auld and Sidhu 2005). Changes in macro-economic environment could also explain why returns may vary: Heckman and Li (2004) claim that economic reforms in China have made monetary returns increase in the last 30 years. Eventually, variations in returns may be driven by unobserved individual characteristics, rather than factors evenly applying to everybody. Returns would then depend on one's ability, taste for knowledge, etc. Testing this assumption, Maier et al. (2004) find that monetary returns vary from negative returns for 20% of the population to above 15% for 25% of them.

Our identification strategy provides a piece of answer to a possible heterogeneity of returns to education: are returns the same if one is forced to stay at school (e.g. via a raising of school leaving age) and if one willingly goes on studying? Angrist and Imbens (1994) point out the importance of knowing what part of the population is affected by the instrument, since the causal effect, which they call *Local Average Treatment Effect*, is estimated solely on that part of the population. Our instrument prevented some individuals from studying until they reached the level of education they would have optimally chosen in the absence of war. On the contrary, instruments like changes in compulsory education laws induce students to attend school longer than they wish. Therefore we estimate returns to education on a specific subpopulation, which is different from the subpopulation affected by compulsory education laws. The most obvious dimension distinguishing them is their motivation to attend school in their last years of schooling. It seems likely that those interested in studying benefit more from an extra year of education, both in terms of acquired knowledge and cognitive skills. Hence motivation would affect the magnitude of human capital improvements due to education. As a consequence, it may be the case that (both monetary and non monetary) returns to education depend on students' motivation in classrooms. If motivation indeed mattered, we should find a larger causal impact than in other studies because our instrument prevented motivated students from going to school whereas compulsory education laws have an impact on children who go beyond their preferred level of education.

Several reasons lead us to believe that Albouy and Lequien (2009) is a relevant benchmark for such a comparison. They work on the same country, France, with the same dataset, and they also measure health with survival rates at a given age. Survival rates are computed at respectively 80 and 50 years old, whereas our dependent variable is survival at 60. Using a regression discontinuity design, they exploit two successive raises in minimum school leaving age for identification, 1923 and 1953 being the first birth cohorts affected by these legislative changes[10]. Albouy and Lequien (2009) rely on compliers of these reforms for identification and observe no significant impact of education on survival rates. Besides compliers' motivation, several other reasons could explain wthy their results differ from ours. For instance, returns to education might depend on the age at which survival is measured. However this age dependence should have a very specific and unlikely shape between 50 and 80 in order to account for significant returns at 60 years old, and for none at 50 and 80. Sudden improvements in medicine could also create generation effects in the health equation. But a similar argument holds, since individuals affected by our instrument were born after the compliers of the first reform, and before the compliers of the second reform.

Cutler and Lleras-Muney (2008) provide a list of possible ways for education to

[10]See footnote 5 for a description of these reforms.

causally affect health; most of them are associated with changes in human capital caused by education (greater income, safer work environment, etc.). Most individuals affected by our instrument were 14 years old in their last year of schooling, which is about compliers' school leaving age in Albouy and Lequien (2009). As contents of courses and teaching methods did not dramatically change between birth cohorts 1923 and 1953, compliers in both studies should have acquired about the same amount of knowledge and skills at school, should motivation not matter. Therefore, any causal path from education to health passing through improvements in human capital should bring out similar results in both studies, as long as motivation is not involved. Moreover Cutler and Lleras-Muney (2008) argue that education could lower one's discount rate, and thus increase one's investment in health. Educated individuals could also have larger social networks, which would positively affect their health. However, these last two causal paths should be triggered, if they indeed existed, both with education laws and war as instruments. So they cannot explain why Albouy and Lequien (2009) and the present study end on opposite conclusions. Eventually, comparing their results with ours suggests that differences in motivation may be a source of heterogeneity in returns to education on health.

5.7 Conclusion

We find a positive and significant impact of education on health. For those affected by our instrument, the probability of being alive at 60 increases by 6 percentage points with each extra year of schooling. Moreover, the precision of our estimations allows us to reject the null hypothesis of exogeneity of education in the survival equation.

According to Albouy and Lequien (2009), raising minimum school leaving age in France did not have any significant causal impact on compliers' survival rate. We claim that this divergence of results is mainly driven by the fact that we estimate returns to education on a specific subpopulation which differs from their set of compliers. Our instrument induced some students to give up school earlier than in the absence of war, whereas compulsory education laws forced some students to exceed their otherwise optimal level of education. Hence our set of compliers is composed of more motivated students during their last years of schooling. Motivation could affect the magnitude of human capital improvements caused by each year of education, and in turn affect returns to education on health.

Conclusion générale

Cette thèse présente 5 essais d'évaluation de politique publique. Nous étudions tout d'abord l'efficacité d'un programme d'accompagnement personnalisé vers l'emploi proposé à des allocataires du Revenu Minimum d'Insertion, et montrons que son coût est supérieur aux économies réalisées sur les versements de l'allocation RMI. Les analyses proposées mettent ensuite en lumière que la durée d'une interruption de carrière affecte négativement la trajectoire salariale des femmes après une naissance. Enfin, elles apportent un début d'explication à la formation de rendements de l'éducation sur la santé hétérogènes au sein de la population : la motivation des élèves à poursuivre leurs études augmenterait l'effet protecteur de l'éducation sur la santé.

Bien qu'elle contienne des résultats originaux, cette thèse ne prétend cependant pas proposer de réponse définitive. Au contraire, son apport principal est à nos yeux la mise en évidence de la difficulté à évaluer rigoureusement un programme et à interpréter ensuite les résultats obtenus.

Nous rappellerons d'abord les principaux résultats de cette thèse, avant d'indiquer leurs limites et d'esquisser les perspectives pour prolonger ces travaux.

Accompagnement d'allocataires du Revenu Minimum d'Insertion

Le chapitre 1 propose l'évaluation d'une politique d'aide à la recherche d'emploi dirigée vers une population particulièrement défavorisée : les personnes qui sont allocataires du Revenu Minimum d'Insertion depuis au moins deux ans. La mise en place d'un protocole d'affectation aléatoire pour déterminer quels allocataires éligibles peuvent bénéficier de l'accompagnement permet d'identifier l'impact de ce programme sur le retour à l'emploi et les montants de RMI versés.

Cette méthodologie innovante est très souple, puisqu'elle autorise certaines personnes du groupe de traitement à refuser de se faire accompagner, et certaines personnes du groupe de contrôle à bénéficier de l'accompagnement. L'hypothèse identifiante est que la proportion de traités doit être plus importante dans le groupe de traitement que dans le groupe de contrôle. C'est le cas ici, même si la différence de taux de participation entre les deux groupes est limitée (17% contre 6%). Cela induit une imprécision dans les estimations, partiellement compensée par la grande taille de l'échantillon (18 000 personnes). Nous ne mettons pas en évidence d'effet causal significatif sur l'emploi, mais le programme réduit cependant le montant des transferts de RMI durant le troisième trimestre de l'expérimentation. Cet impact est accentué sur certaines sous-populations, dont les personnes allocataires du RMI depuis moins de 4 ans lors de leur entrée dans l'expérimentation.

Ce programme a un coût direct pour le Conseil Général qui l'a mis en place, puisqu'il doit rétribuer le prestataire pour l'accompagnement effectué. Ce coût direct est supérieur aux économies réalisées sur les allocations RMI versées grâce au programme. Le programme d'accompagnement est donc coûteux pour le budget du département.

Une originalité de ce chapitre est que l'entrée dans le traitement ne s'effectue pas forcément juste après le tirage au sort. Les allocataires du RMI débutent en effet leur accompagnement au fil des mois qui suivent leur affectation. Cette dynamique a des implications fortes sur les paramètres qu'il est possible d'identifier. Deux cas de figure sont possibles : un LATE peut être estimé lorsque l'impact causal dépend de la durée en traitement, ou lorsque l'impact causal est corrélé avec la probabilité d'entrée en traitement. Le modèle n'est cependant pas identifié lorsque ces deux hypothèses sont posées simultanément. Nous avons privilégié dans nos commentaires la première hypothèse, car c'est celle qui semble la plus réaliste. Il serait cependant utile de pouvoir autoriser l'impact causal à varier suivant ces deux dimensions à la fois.

Une limite de notre étude est que nous évaluons la prestation d'accompagnement dans son ensemble, et ne pouvons pas nous prononcer sur l'efficacité de chacune des composantes de l'action du prestataire (bilan de compétences, définition du projet professionnel, atelier d'écriture de CV, démarchage d'entreprises pour rechercher des offres d'emploi, etc.). Il aurait fallu pour cela pouvoir affecter aléatoirement les allocataires à chacune de ces actions, ce qui n'a pas été possible.

De plus, le processus allant du tirage au sort jusqu'à l'entrée effective en accompagnement des allocataires n'a pas eu l'efficacité espérée. Le prestataire a essayé de contacter les personnes éligibles par courrier ou téléphone. Une fois contactées, ces personnes étaient invitées à une réunion collective d'information, à l'issue de laquelle elles décidaient d'accepter ou non d'entrer dans le programme. Il était clair dès le départ que la population ciblée était particulièrement éloignée du marché du travail, et qu'il serait difficile de convaincre la totalité du groupe de traitement de suivre l'accompagnement. Le faible taux de participation constaté dans le groupe de traitement (17%) suggère cependant que des démarches plus efficaces auraient pu être entreprises (démarchage à domicile, stand mobile dans les lieux susceptibles d'accueillir beaucoup de personnes éligibles, etc.). Le mode de rémunération proposé au prestataire, fixé par le Conseil général, aurait également pu contenir une incitation plus forte à faire adhérer au programme.

Une autre limite est le contenu de la variable d'emploi que nous utilisons. Elle provient de fichiers de la caisse d'allocations familiales du département concerné, et nous renseigne mensuellement sur la situation professionnelle des allocataires. Nous savons donc si les personnes sont en emploi ou non, avec suivant le cas le type d'emploi (CDD, contrat aidé, etc.) ou de non emploi (sans activité, chômeur non indemnisé, etc.). Lorsque l'allocataire est en emploi, nous ne connaissons cependant pas le temps de travail, le salaire, l'employeur, la durée du contrat, etc. Etant donné que le public cible est très défavorisé, il aurait été intéressant de savoir quel type d'emploi les personnes du groupe de traitement retrouvent, et s'il s'agit d'emplois similaires à ceux trouvés par les personnes du groupe de contrôle.

Enfin, le contexte particulier de cette expérimentation (un public très ciblé, à une époque donnée, dans un département donné, avec un prestataire donné) ne permet pas de généraliser les résultats. Il faudrait par exemple répliquer ce dispositif à l'identique dans d'autres territoires pour savoir si les résultats obtenus dans un département sont liés aux caractéristiques de ce département, ou au contraire sont généralisables à l'ensemble de la France. Ce constat n'est cependant pas propre à ce chapitre, et se retrouve dans beaucoup d'autres études.

L'impact de la durée d'un arrêt de carrière sur les salaires

Les chapitres 2 et 3 s'intéressent aux variations de salaire provoquées par une interruption de carrière suite à une naissance. Ils montrent que la durée d'un arrêt de carrière semble avoir un impact négatif sur les salaires une fois que le retour sur le marché du travail est effectué.

La réforme de l'Allocation Parentale d'Education de 1994 permet l'identification de cet impact causal sur les femmes qui ont été affectées par la réforme. Une estimation non paramétrique par doubles et triples différences est proposée au chapitre 2, tandis que le chapitre 3 présente l'estimation d'un modèle structurel. Les résultats produits par ces deux méthodes sont qualitativement similaires : la durée d'une interruption de carrière suite à une naissance a un impact causal négatif sur les salaires. Si les ordres de grandeur de l'effet causal sont les mêmes dans les deux chapitres, le modèle structurel donne des estimations plus précises.

Ces deux méthodes sont fondées sur des hypothèses différentes. En particulier, le modèle structurel est a priori sujet à une éventuelle mauvaise spécification des équations de salaire et de sélection. La comparabilité des résultats avec ces deux méthodes indique qu'une telle mauvaise spécification, si elle existe, ne biaise pas de manière importante les estimations.

La principale limite de ces deux chapitres est qu'ils étudient le salaire lors du retour à l'emploi, mais aucune des autres dimensions de l'emploi retrouvé. Il est tout à fait possible que les mères terminant une période d'APE retrouvent certes un emploi, mais aient des difficultés à retrouver un emploi de qualité en termes de temps de travail, intérêt du poste, etc. Les DADS ne permettent malheureusement d'approcher cet aspect du retour à l'emploi que par le salaire perçu. Le chapitre 2 propose une discussion sur l'effet que pourrait avoir la prise en compte du temps de travail sur nos estimations et leur interprétation (si par exemple les mères reviennent sur le marché du travail systématiquement sur des emplois à temps partiel). Mais il ne peut cependant donner que des pistes d'explication faute de données disponibles. Il y a là un prolongement important à donner à ces deux chapitres.

Cette thèse met en évidence que l'impact négatif sur le salaire d'une interruption de carrière existe, et semble durer plusieurs années après le retour à l'emploi. Il est donc possible que la durée d'interruption continue à avoir un impact sur les revenus après la fin de la vie active, puisque le montant de la pension de retraite dépend directement des salaires perçus durant les 25 meilleures années de carrière dans le secteur privé. Une extension intéressante consisterait à quantifier les conséquences de cette baisse de salaire sur le montant de la pension de retraite des mères, et à déterminer si elles sont compensées par les différents droits familiaux qui existent dans le calcul des pensions.

Depuis 2004, le Complément de Libre Choix d'Activité a remplacé l'APE. Ce dispositif est resté très similaire à l'APE, mais il a étendu l'éligibilité aux parents d'un enfant de rang 1, l'allocation étant versée pendant une durée limitée à 6 mois (au lieu de 3 ans pour les naissances de rang supérieur ou égal à 2). Un prolongement possible serait de regarder quel impact le CLCA a eu sur la trajectoire salariale des mères d'un premier enfant. En particulier, le fait qu'il soit limité à 6 mois devrait permettre de déterminer si des non linéarités existent dans le lien entre la durée d'interruption et le salaire.

Enfin, le Parlement européen a adopté le 20 octobre 2010 une proposition visant à augmenter la durée minimale du congé maternité en Europe, la faisant ainsi passer de 14 semaines à 20 semaines (le congé maternité dure actuellement 16 semaines en France pour une première naissance). Si elle est acceptée et mise en place par les Etats membres, cette proposition permettrait à toutes les mères qui travaillent de disposer d'un congé maternité intégralement payé de 20 semaines. Cette proposition est présentée comme une avancée sociale par les parlementaires européens, mais il est probable qu'elle ait également des répercussions négatives sur le déroulement de la carrière des mères, et en particulier sur leur salaire. Cet exemple illustre le besoin de documenter les mécanismes qui déterminent les trajectoires salariales, afin que nos dirigeants politiques disposent des informations nécessaires pour prendre leurs décisions. Cette thèse apporte un éclairage sur l'un d'entre eux (les arrêts consécutifs à une naissance), mais beaucoup reste à faire.

Les rendements de l'éducation sur la santé

Les deux derniers chapitres de cette thèse estiment des rendements de l'éducation sur la santé.

Le chapitre 4 exploite les discontinuités créées par deux allongements successifs de l'âge de fin de scolarité obligatoire. Ce type de réforme a été utilisé dans de nombreux autres pays pour identifier des rendements de l'éducation (voir par exemple Lleras-Muney 2005, Oreopoulos 2006), le plus souvent dans le cadre classique des variables instrumentales. Nous mettons ici en œuvre des régressions par discontinuités, car nous pensons qu'il s'agit d'une modélisation particulièrement adaptée à ce contexte.

Le chapitre 5 utilise une autre source d'identification, plus originale que la précédente. La seconde guerre mondiale a perturbé (entre autres choses) le système éducatif français, ainsi que les priorités des personnes vivant à cette époque. Nous constatons que l'âge de fin d'études moyen des générations qui étaient adolescentes pendant la guerre est plus faible que d'habitude, et montrons que cela peut servir à identifier les rendements de l'éducation.

Les estimations dans ces deux chapitres donnent à première vue des résultats contradictoires : les *compliers* du chapitre 4 n'ont pas vu leur santé s'améliorer grâce à une scolarité plus longue, tandis que la santé des *compliers* du chapitre 5 s'est dégradée à cause d'une scolarité plus courte. Cependant, la nature même de ces chocs exogènes fait que les estimations sont identifiées sur des individus très différents. En particulier, ces *compliers* diffèrent par leur motivation à poursuivre leurs études dans leurs dernières années de scolarité. Les premiers ont dû prolonger leurs études plus longtemps qu'ils ne le désiraient, alors que les autres ont au contraire dû abréger leur scolarité contre leur volonté. Cela suggère que les rendements de l'éducation pourraient varier suivant les individus, et qu'une dimension importante de cette hétérogénéité serait la motivation à poursuivre les études. Cette possible hétérogénéité n'est à notre connaissance pas mentionnée dans la littérature, la plupart des auteurs se concentrant sur des rendements pouvant varier avec l'âge de fin d'études ou le type de diplôme.

Les travaux contenus dans cette thèse apportent donc des éléments permettant de juger de l'ampleur des effets engendrés par la mise en place d'une politique publique ou la survenue d'un événement ayant eu des répercutions sur le société française. Comme souvent dans un travail de recherche de ce type, nous proposons des résultats originaux, tout en constatant l'étendue de ce qu'il reste à faire...

Bibliographie

Aakvik, A., Salvanes, K. and Vaage, K. (2010), 'Measuring heterogeneity in the returns to education in norway using educational reforms', *European Economic Review* **54**(3), 483–500. 174

Abbring, J. and van den Berg, G. (2003), 'The non parametric identification of treatment effects in duration models', *Econometrica* **71**(5), 1491–1517. 29

Abbring, J. and van den Berg, G. (2005), 'Social experiments and instrumental variables with duration outcomes', *IFS Working Paper* **05/19**. 20, 34, 35

Adams, S. (2002), 'Educational attainment and health : Evidence from a sample of older adults', *Education Economics* **10**(1), 97–109. 122, 123, 124, 144, 153

Albouy, V. and Lequien, L. (2009), 'Does compulsory education lower mortality ?', *Journal of Health Economics* **28**(1), 155–168. 153, 154, 157, 175, 176

Albrecht, J., Edin, P.-A., Sundström, M. and Vroman, S. (1999), 'Career interruptions and subsequent earnings : A reexamination using Swedish data', *The Journal of Human Resources* **34**(2), 294–311. 69, 110

Angrist, J. (1990), 'Lifetime earnings and the Vietnam era draft lottery : Evidence from Social Security administrative records', *The American Economic Review* **80**(3), 313–336. 80, 168, 169

Angrist, J. and Imbens, G. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475. 7, 8, 14, 20, 31, 168, 174

Angrist, J. and Imbens, G. (1995), 'Two-stage least squares estimation of average causal effects in models with variable treatment intensity', *Journal of the American Statistical Association* **90**, 431–442. 132

Angrist, J., Imbens, G. and Rubin, D. (1996), 'Identification of causal effects using instrumental variables', *Journal of the American Statistical Association* **91**(434), 444–455. 75, 83, 130

Angrist, J. and Krueger, A. (1994), 'Why do World War II veterans earn more than nonveterans ?', *Journal of Labor Economics* **12**(1), 74–97. 169

Angrist, J. and Lavy, V. (1999), 'Using Maimonides' rule to estimate the effect of class size on scholastic achievement', *Quarterly Journal of Economics* **114**(2), 533–575. 11, 128

Arendt, J. (2005), 'Education effects on health : A panel data analysis using school reform for identification', *Economics of Education Review* **24**(2), 149–160. 12, 122, 123, 126, 144, 152, 153

Arkes, J. (2003), 'Does schooling improve adult health ?', *RAND Working Paper* **DRU-3051**. 122, 123, 125, 145, 153

Auld, M. and Sidhu, N. (2005), 'Schooling, cognitive ability, and health', *Health Economics* **14**, 1019–1034. 123, 126, 153, 174

Banerjee, A. and Duflo, E. (2008), 'The experimental approach to development economics', *CEPR Discussion Paper* **7037**. 6

Bardaji, J., Sédillot, B. and Walreat, E. (2003), 'Un outil de prospective des retraites : le modèle de microsimulation Destinie', *Economie et Prévision* **160-161**. 69, 110

Bayet, A. (1997), 'Continuous careers, incomplete careers, and wages', *Insee Studies* **7**. 109

Beblo, M., Bender, S. and Wolf, E. (2006), 'The wage effects of entering motherhood', *Discussion Paper* **06-053**. 50

Bedard, K. and Deschenes, O. (2006), 'The long-term impact of military service on health : Evidence from World War II and Korean war veterans', *American Economic Review* **96**(1), 176–194. 169

Behaghel, L., Crépon, B. and Gurgand, M. (2009), 'Evaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privés de placement et le programme Cap vers l'entreprise', *Research report, Paris* . 18, 19

Belzil, C. (2007), 'The return to schooling in structural dynamic models : A survey', *European Economic Review* **51**, 1059–1105. 164

Benabou, R., Kramarz, F. and Prost, C. (2009), 'The French Zones d'Education Prioritaires : Much ado about nothing ?', *Economics of Education Review* **28**(3), 354–356. 1

Berger, M. and Leigh, J. (1989), 'Schooling, self-selection, and health', *The Journal of Human Resources* **24**(3), 433–45. 122, 124, 145

Berman, E., Iannaccone, R. and Ragusa, G. (2006), 'From empty pews to empty cradles : Fertility decline among european catholics', *University of San Diego, mimeo*. 80

Blanchet, D. and Pennec, S. (1997), 'Is the rise in female participation linked to fertility rate trends ?', *Insee Studies* **9**. 98

Blundell, R. (2006), 'Earned income tax credit policies : Impact and optimality, the adam smith lecture, 2005', *Labour Economics* **13**, 423–443. 18

Blundell, R., Dias, M. C., Meghir, C. and van Reenen, J. (2004), 'Evaluating the employment impact of a mandatory job search program', *Journal of the European Economic Association* **2**, 569–606. 18

Bonnal, L., Fougère, D. and Sérandon, A. (1997), 'Evaluating the impact of french employment policies on individual labour market histories', *Review of Economic Studies* **6**, 683–713. 1, 18

Boone, J., Fredriksson, P., Holmlund, B. and van Ours, J. (2007), 'Optimal unemployment insurance with monitoring and sanctions', *Economic Journal* **117**, 399–421. 18

Bopp, M. and Minder, C. E. (2003), 'Mortality by education in German speaking Switzerland 1990-1997 : Results from the Swiss National Cohort', *International Journal of Epidemiology* **32**, 346–354. 122

Brouillette, D. and Lacroix, G. (2010), 'Heterogeneous treatment and self-selection in a wage subsidy experiment', *Journal of Public Economics* **94**(7-8), 479–492. 18

Buding, M. J. and England, P. (2001), 'The wage penalty of motherhood', *American Sociological Review* **66**(2), 204–225. 50, 81

Card, D. (2001), 'Estimating the return to schooling : Progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–1160. 164

Card, D. and Hyslop, D. (2005), 'Estimating the effects of a time-limited earnings subsidy for welfare leavers', *Econometrica* **73**. 18, 29

Card, D., Kluve, J. and Weber, A. (2010), 'Active labour market policy evaluations : A meta-analysis', *The Economic Journal* **120**(548), 452–477. 18

Card, D., Michalopoulos, C. and Robins, P. (2005), 'When financial incentives pay for themselves : Evidence from a randomized social experiment for welfare recipients', *Journal of Public Economics* **89**. 18

Card, D. and Robins, P. (2005), 'How important are entry effects in financial incentive programs for welfare recipients ? Experimental evidence from the self-sufficiency project', *Journal of Econometrics* **125**. 18

Case, A., Fertig, A. and Paxson, C. (2005), 'The lasting impact of childhood health and circumstance', *Journal of Health Economics* **24**, 365–389. 170

Chamberlain, G. (1980), 'Analysis of covariance with qualitative data', *Review of Economic Studies* **47**, 225–238. 91

Chen, K. and Lange, F. (2008), 'Education, information, and improved health : Evidence from breast cancer screening', *IZA Discussion Paper* **3548**. 174

Chen, Y. and Zhou, L.-I. (2007), 'The long term health and economic consequences of the 1959-1961 famine in China', *Journal of Health Economics* **26**, 659–681. 166, 170, 171

Chou, S.-Y., Grossman, M. and Saffer, H. (2004), 'An economic analysis of adult obesity : Results from the behavioral risk factor surveillance system', *Journal of Health Economics* **23**, 565–587. 122

Cipollone, P., Radicchia, D. and Rosolia, A. (2006), 'The effect of education on youth mortality', *Bank of Italy, mimeo.* . 127, 128, 153

Clark, D. and Royer, H. (2008), 'The effect of education on adult mortality and health : Evidence from the united kingdom', *mimeo.* . 153

Costa, D. (2003), 'Understanding mid-life and older age mortality declines : Evidence from Union Army veterans', *Journal of econometrics* **112**, 175–192. 170

Crépon, B., Dejemeppe, M. and Gurgand, M. (2005), 'Counceling the unemployed : Does it lower unemployment duration and recurrence ?', *IZA Discussion Paper* **1796**. 18

Cutler, D. and Lleras-Muney, A. (2008), Education and health : Evaluating theories and evidence, *in* 'Making Americans Healthier : Social and Economic Policy as Health Policy', Russell Sage Foundation, chapter 2, pp. 30–60. 153, 175, 176

Cutler, D., Miller, G. and Norton, D. (2007), 'Evidence on early-life income and late-life health from America's Dust Bowl era', *Proceedings of the National Academy of Science* **104**(33), 13244–13249. 170

Davies, R. and Pierre, G. (2005), 'The family gap in pay in europe : A cross-country study', *Labour Economics* **12**(4), 469–486. 50

De Walque, D. (2003), How Does Education Affect Health Decisions ? The Cases of Smoking and HIV/AIDS, PhD thesis, University of Chicago, Department of Economics. 122, 125

De Walque, D. (2004), 'Education, information, and smoking decisions, evidence from smoking histories 1940-2000', *World Bank Policy Research Working Paper* **3362**. 121, 122, 125, 174

De Walque, D. (2007), 'Does education affect smoking behaviors ? Evidence using the Vietnam draft as an instrument for college education', *Journal of Health Economics* **26**(5), 877–895. 122, 127, 153, 168

Deaton, A. (2008), 'Randomization in the tropics, and the search for the elusive keys to economic development', *The Keynes Lecture, British Academy* . 6

Decker, P. T. and O'Leary, C. J. (1994), 'Evaluating pooled evidence from the reemployment bonus experiments', *Upjohn Institute Working Paper* **94-28**. 18

Del Boca, D. and Wetzels, C. (2008), *Social Policies, Labour Markets and Motherhood : A Comparative Analysis of European Countries*, Cambridge University Press, New York. 50, 80

Dolton, P. and O'Neill, D. (1996), 'Unemployment duration and the restart effect : some experimental evidence', *Economic Journal* **106**, 387–400. 18

Dolton, P. and O'Neill, D. (2002), 'The long-run effects of unemployment monitoring and work-search programs : Experimental evidence from the United Kingdom', *Journal of Labor Economics* **20**, 381–403. 18

Doyle, O., Harmon, C. and Walker, I. (2005), 'The impact of parental income and education on the health of their children', *IZA Discussion Paper* **1832**. 12, 122, 126, 144

Duncan, T. (1997), 'Sailors, scurvy and science', *Journal of the Royal Society of Medicine* **90**(6), 50–54. 6

Dustmann, C. and Rochina-Barrachina, M. E. (2007), 'Selection correction in panel data models : An application to the estimation of females' wage equations', *Econometrics Journal* **10**, 263–293. 87, 88

Elo, I. and Preston, S. (1996), 'Educational differentials in mortality : United States, 1975-1985', *Social Science and Medicine* **42**(1), 47–57. 122, 156

Felfe, C. (2006), 'The child penalty : A compensating wage differential', *ENEPRI Research Report* **22**. 50, 99, 109

Fernandez, R. and Fogli, A. (2006), 'Fertility : The role of culture and family experience', *Journal of the European Economic Association* **4**(2-3), 552–561. 80

Fisher, R. A. (1935), *Design of Experiments*, Oliver and Boyd, Edinburgh. 6

Fougère, D., Kamionka, T. and Priéto, A. (2010), 'L'éfficacité des mesures d'accompagnement sur le retour à l'emploi', *Revue Economique* **61**, 599–612. 1, 18

Francesconi, M. (2002), 'A joint dynamic model of fertility and work of married women', *Journal of Labor Economics* **20**(2), 336–380. 81

Fuchs, V. R. (1982), *Time Preference and Health : An Exploratory Study*, The University of Chicago Press, pp. 93–120. 11

Gerfin, M. and Lechner, M. (2002), "microeconometric evaluation of the active labour market policy in Switzerland', *Economic Journal* **112**, 854–893. 18

Glied, S. and Lleras-Muney, A. (2003), 'Health inequality, education and medical innovation', *NBER Working Paper* **9738**. 12, 121, 125, 128, 144, 145

Glied, S. and Lleras-Muney, A. (2008), 'Technological innovation and inequality in health', *Demography* **45**(3), 741–761. 153, 174

Goldman, D. and Smith, J. (2002), 'Can patient self-management help explain the SES health gradient ?', *Proceedings of the National Academy of Science* **99**(16), 10929–10934. 120, 174

Grenet, J. (2003), Suffit-il d'allonger la durée de scolarité obligatoire pour augmenter les salaires ?, Master's thesis, EHESS, Paris. 147

Gritz, M. (1993), 'The impact of training on the frequency and duration of employment', *Journal of Econometrics* **57**, 21–51. 1, 18

Groot, W. and Maassen van den Brink, H. (2007), 'The health effects of education', *Economics of Education Review* **26**, 186–200. 122, 123, 127

Grossman, M. (1972), 'On the concept of health capital and the demand for health', *The Journal of Political Economy* **80**(2), 223–255. 11, 120, 146, 152

Grossman, M. (2004), 'The demand for health, 30 years later : A very personal retrospective and prospective reflection', *Journal of Health Economics* **23**, 629–636. 11, 122, 153, 164

Gupta, N. and Smith, N. (2002), 'Children and career interruptions : The family gap in denmark', *Economica* **69**, 609–629. 50

Hahn, J., Todd, P. and van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression discontinuity design', *Econometrica* **69**, 201–209. 12, 128, 129, 130

Heckman, J. (1976), 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models', *The Annals of Economic and Social Measurement* **5**, 475–492. 10, 87

Heckman, J. and Li, X. (2004), 'Selection bias, comparative advantage and heterogeneous returns to education : Evidence from china in 2000', *Pacific Economic Review* **9**(3), 155–171. 174

Hotchkiss, J. and Pitts, M. (2007), 'The role of labour market intermittency in explaining gender wage differentials', *American Economic Review* **97**(2), 417–421. 50

Ichino, A. and Winter-Ebmer, R. (2004), 'The long-run educational cost of World War II', *Journal of Labor Economics* **22**(1), 57–86. 154, 159, 164, 171

Insee (2007), 'La situation démographique en 2005 - Mouvement de la population', *Insee Résultats* **66**. 165, 166

Jäckle, R. (2007), 'Health and wages : Panel data estimates considering selection and endogeneity', *Ifo Working Paper* **43**. 88

Jusot, F. (2003), Revenu et Mortalité : Analyse Économique des Inégalités Sociales de Santé en France, PhD thesis, EHESS, Paris. 121

Kamionka, T. and Lacroix, G. (2008), 'Assessing the external validity of an experimental wage subsidy', *Annales d'Économie et de Statistique* **91-92**. 18

Kannisto, V., Christensen, K. and Vaupel, J. (1997), 'No increased mortality in later life for cohorts born during famine', *American Journal of Epidemiology* **145**(11), 987–994. 165, 170

Katz, L. F., Kling, J. R. and Liebman, J. B. (2001), 'Moving to opportunity in Boston : Early results of a randomized mobility experiment', *The Quarterly Journal of Economics* **116**(2), 607–654. 121

Keane, M. and Wolpin, K. (2002*a*), 'Estimating welfare effects consistent with forward-looking behaviour, I : Lessons from a simulation exercise', *Journal of Human Resources* **37**(3), 570–599. 81

Keane, M. and Wolpin, K. (2002*b*), 'Estimating welfare effects consistent with forward-looking behaviour, II : Empirical results', *Journal of Human Resources* **37**(3), 600–622. 81

Kenkel, D. S. (1991), 'Health behavior, health knowledge, and schooling', *The Journal of Political Economy* **99**(2), 287–305. 122, 124

Kenkel, D. S., Lillard, D. and Mathios, A. (2006), 'The roles of high school completion and GED receipt on smoking and obesity', *Journal of Labor Economics* **24**(3), 635–660. 122, 126

Koubi, M. (2003), 'Les carrières salariales par cohorte de 1967 à 2000', *Economie et Statistique* **369-370**. 69, 110

Kremer, M. and Miguel, E. (2004), 'Worms : Identifying impacts on education and health in the presence of treatment externalities', *Econometrica* **72**(1), 159–217. 152

Kunze, A. and Ejrnaes, M. (2004), 'Wage dips and drops around first birth', *IZA Discussion Paper* **1011**. 50

Kyriazidou, E. (1997), 'Estimation of a panel data sample selection model', *Econometrica* **65**, 1135–1164. 87

Lalive, R. and Zweimüller, J. (2005), 'Does parental leave affect fertility and return-to-work ? Evidence from a true natural experiment', *IZA Discussion Paper* **1613**. 50

Laroque, G. and Salanié, B. (2008), 'Does fertility respond to financial incentives ?', *CESifo Working Paper* **2339**. 81

Legendre, F., Lorgnet, J.-P., Mahieu, R. and Thibault, F. (2003), 'Etat des lieux des prestations petite enfance avant la mise en place de la prestation d'accueil du jeune enfant', *CNAF L'e-ssentiel* **16**. 52

Lind, J. (1753), *A Treatise of the Scurvy*, Sands, Murrav and Cochran, Edinburgh. 6

Lleras-Muney, A. (2005), 'The relationship between education and adult mortality in the United States', **72**, 189–221. 12, 122, 126, 128, 143, 144, 145, 148, 153, 184

Maier, M., Pfeiffer, F. and Pohlmeier, W. (2004), 'Returns to education and individual heterogeneity', *ZEW Discussion Paper* **04-34**. 174

Michael, R. T. (1973), 'Education in nonmarket production', *The Journal of Political Economy* **81**(2), 306–327. 120

Milligan, K., Moretti, E. and Oreopoulos, P. (2004), 'Does education improve citizenship ? Evidence from the United States and the United Kingdom', *Journal of Public Economics* **88**, 1667–1695. 120

Mundlak, Y. (1978), 'On the pooling of time series and cross section data', *Econometrica* **46**, 69–85. 87, 91, 92, 94

Ondrich, J., Spiess, C. and Yang, Q. (2002), 'The effect of maternity leave on women's pay in Germany 1984-1994', *DIW Discussion Paper* **289**. 50

Oreopoulos, P. (2006), 'Estimating average and local average treatment effects of education when compulsory schooling laws really matter', *The American Economic Review* **96**(1), 152–175. 12, 122, 123, 127, 144, 145, 152, 153, 184

Oreopoulos, P. (2007), 'Do dropouts drop out too soon ? Wealth, health and happiness from compulsory schooling', *Journal Of Public Economics* **91**(11-12), 2213–2229. 153

Oreopoulos, P. and Salvanes, K. (2009), 'How large are returns to schooling ? hint : Money isn't everything', *NBER Working Paper* **15339**. 174

Pailhé, A. and Solaz, A. (2006), 'Vie professionnelle et naissance : la charge de la conciliation repose essentiellement sur les femmes', *Population et Sociétés* **436**. 8, 51, 53, 55, 76, 85, 99

Painter, R., Roseboom, T., Bossuyt, P., Osmond, C., Barker, D. and Bleker, O. (2005), 'Adult mortality at age 57 after prenatal exposure to the Dutch famine', *European Journal of Epidemiology* **20**(8), 673–676. 170

Paxton, R. (1982), *Vichy France : Old Guard and New Order, 1940-1944*, Columbia University Press, New York. 169

Piketty, T. (2005), 'Impact de l'allocation parentale d'education sur l'activité féminine et la fécondité en France', *Histoires de familles, histoires familiales* **156**, 79–109. 8, 50, 51, 53, 55, 58, 61, 63, 76, 81, 82, 85, 96, 98, 101

Piketty, T. and Valdenaire, M. (2006), 'L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français : Estimations à partir du panel primaire 1997 et du panel secondaire 1995', *Ministère de l'éducation nationale, Les Dossiers* **173**. 1

Pison, G. (2009), 'France 2008 : Pourquoi le nombre de naissances continue-t-il d'augmenter ?', *Population et Sociétés* **454**. 63

Pouget, J., Rapoport, B. and Serravalle, S. (2007), 'The impact of unemployment duration on wages : Evidence from French panel data 1984-2001', *mimeo*. 88

Prost, A. (1968), *Histoire de l'enseignement en France, 1800-1967*, Editions Armand Colin. 136

Rochina-Barrachina, M. E. (1999), 'A new estimator for panel data sample selection models', *Annales d'Economie et Statistique* **55/56**, 153–181. 87

Roseboom, T., van der Meulen, J., Osmond, C., Barker, D., Ravelli, C. and Bleker, O. (2001), 'Adult survival after prenatal exposure to the Dutch famine 1944-45', *Paediatric and Perinatal Epidemiology* **15**(3), 220–225. 165, 170

Rosen, S. and Willis, J. R. (1979), 'Education and self-selection', *Journal of Political Economy* **87**, 7–36. 132

Ross, C. and Mirowsky, J. (1999), 'Refining the association between education and health : The effects of quantity, credential, and selectivity', *Demography* **36**(4), 445–460. 147

Rubin, D. (1974), 'Estimating causal effects of treatments in randomized and non-randomized studies', *Journal of Educational Psychology* **66**(6), 688–701. 3

Ruhm, C. (1998), 'The economic consequences of parental leave mandates : Lessons from Europe', *The Quarterly Journal of Economics* **113**(1), 285–317. 50

Semykina, A. and Wooldridge, J. (2005), 'Estimating panel data models in the presence of endogeneity and selection : Theory and application', *mimeo*. 112, 115

Semykina, A. and Wooldridge, J. (2010), 'Estimating panel data models in the presence of endogeneity and selection', *Journal of Econometrics* **157**(2), 375–380. 10, 88, 89, 90, 91

Sianesi, B. (2004), 'An evaluation of the Swedish system of active labor market programs in the 1990s', *Review of Economics and Statistics* **86**, 133–155. 18

Silles, M. (2009), 'The causal effect of education on health : Evidence from the united kingdom', *Economics of Education Review* **28**, 122–128. 153, 157, 164

Sirinelli, J.-F., Vandenbussche, R. and Vavasseur-Desperriers, J. (2004), *La France de 1914 à Nos Jours*, Presses Universitaires de France, Paris. 155, 167

Spandorfer, J., Karras, D., Hughes, L. and Caputo, C. (1995), 'Comprehension of discharge instructions by patients in an urban emergency department', *Annals of Emergency Medicine* **25**(1), 71–74. 120

Spasojevic, J. (2003), Effect of Education on Adult Health in Sweden : Results from a Natural Experiment, PhD thesis, City University of New York Graduate Center. 122, 124, 144

Thistlethwaite, D. and Campbell, D. (1960), 'Regression-discontinuity analysis : An alternative to the ex post facto experiment', *Journal of Educational Psychology* **51**, 309–317. 11, 128

van den Berg, G. and van der Klaauw, B. (2006), 'Counseling and monitoring of unemployed workers : Theory and evidence from a controlled social experiment', *International Economic Review* **47**(3), 895–936. 18

van der Klaauw, W. (2002), 'Estimating the effect of financial aid offers on college enrollment : A regression-discontinuity approach', *International Economic Review* **43**(4), 1249–1287. 11, 128, 131, 132

van Kippersluis, H., O'Donnell, O. and van Doorslaer, E. (2009), 'Long run returns to education : Does schooling leads to an extended old age ?', *Tinbergen Institute Discussion Paper* **037/3**. 153, 156

van Oers, J. (2003), *Health on Course ? The 2002 Dutch Public Health Status and Forecasts Report*, The National Institute for Public Health and the Environment, Bilthoven, The Netherlands. 122

Woodbury, S. A. and Spiegelman, R. G. (1987), 'Bonuses to workers and employers to reduce unemployment : randomized trials in Illinois', *American Economic Review* **77**, 513–530. 1, 18

Wooldridge, J. (1995), 'Selection correction for panel data models under conditional mean independence assumptions', *Journal of Econometrics* **68**, 115–132. 87, 88

Wright, G. (1962), 'Reflections on the French Resistance (1940-1944)', *Political Science Quarterly* **77**(3), 336–349. 169

Essays on public policy evaluation,
covering education, health and labor market policies

This thesis provides 5 essays on public policy evaluation. We first consider a job-search counseling program targeted at individuals that have been on welfare for at least two years. We evaluate the impact of this policy on job access and its cost-effectiveness, using a random allocation of welfare recipients to the program. Its effects on employment are, in general, far too small to ensure that the program is cost-effective for the public budget. Then we test whether parental leave duration has a causal impact on later wages, with two modeling strategies : the first one uses a difference in difference approach, whereas the other imposes more structure and specifies a selection equation. Our results suggest that parental leave duration has a negative causal impact on wages. The last two chapters of this thesis are devoted to returns to education on health, with the use of two different instrumental variables. The first instrument induced some students to give up school earlier than they wished, whereas the other instrument forced some students to exceed their otherwise optimal level of education. We find significant returns only in the first case, when the set of compliers is composed of motivated students. Hence this suggests that differences in motivation may be a source of heterogeneity in returns to education on health.

Keywords : evaluation, public policy, labor market, counseling program, wage, parental leave, education, health

Essais d'évaluation de politique publique dans les champs de l'éducation, de la santé et des politiques d'emploi

Cette thèse présente 5 essais d'évaluation de politique publique. Nous étudions tout d'abord l'efficacité d'un programme d'accompagnement personnalisé vers l'emploi proposé à des allocataires du Revenu Minimum d'Insertion, et montrons que son coût est supérieur aux économies réalisées sur les versements de l'allocation. Les analyses proposées mettent ensuite en lumière que la durée d'une interruption de carrière affecte négativement la trajectoire salariale des femmes après une naissance. Enfin, elles apportent un début d'explication à la formation de rendements de l'éducation sur la santé hétérogènes au sein de la population : la motivation des élèves à poursuivre leurs études augmenterait l'effet protecteur de l'éducation sur la santé.

Sur le plan méthodologique, cette thèse met en œuvre 5 modélisations économétriques différentes pour identifier l'impact causal : une expérimentation aléatoire (chapitre 1), des doubles différences (chapitre 2), une instrumentation sur données de panel avec un modèle structurel (chapitre 3), des régressions par discontinuité (chapitre 4), et enfin le cadre classique des variables instrumentales (chapitre 5).

Mots clés : évaluation, politique publique, marché du travail, accompagnement vers l'emploi, salaire, interruption de carrière, éducation, santé