

Salaire, productivité et demande de travailleurs âgés

Patrick Aubert (INSEE et CREST-LEI)¹

VERSION PROVISOIRE

13 février 2003

Dans cette étude, nous estimons le profil de la productivité selon l'âge par l'estimation d'une fonction de production sur une base couvrant plus de 33 000 entreprises et deux millions de salariés suivi de 1994 à 2000. Cette estimation souffre d'un biais d'endogénéité important, dû à la causalité inverse de la productivité sur la structure par âge. Il ressort de nos résultats que la relation négative entre productivité et âge fréquemment estimée lorsque ce biais n'est pas pris en compte ne vient pas du fait que les travailleurs âgés sont moins productifs que les plus jeunes, mais plutôt de celui qu'ils sont plus souvent dans les entreprises moins productives.

Après correction de ce biais, nous estimons une productivité croissante avec l'âge, cette croissance se ralentissant à partir de 40 ans et devenant presque nulle au delà de 50 ans. Ce résultat semble invalider l'idée que la faible demande de travailleurs âgés par les entreprises serait due à une moindre productivité de ces travailleurs.

Nous estimons de manière similaire le profil en fonction de l'âge du coût salarial et du rapport de la productivité sur le salaire. Le manque de robustesse des résultats nous empêche malheureusement de conclure sur l'adéquation entre salaire et productivité.

Mots clés : productivité du travail, demande de travailleurs âgés, fonction de production

Classification JEL : J24, J31

1. Introduction

Une caractéristique frappante de la littérature économique sur l'activité des plus âgés est la dissymétrie qui existe entre le nombre d'études consacrées à l'offre de travail de ces travailleurs, aux incitations qu'ils reçoivent du fait des systèmes de pensions et des perspectives sur le marché du travail et aux conséquences de ces incitations sur la décision de cesser leur activité, et à celui des études consacrées à la demande de travail, c'est à dire aux comportements des entreprises quant à la gestion de leur main d'œuvre âgée. Si la plupart des auteurs soulignent l'importance de cet aspect et la nécessité de le prendre en compte dans toute explication de l'activité ou l'emploi des plus de 50 ans, les estimations empiriques de la demande de ces travailleurs par les entreprises et de ses déterminants sont beaucoup plus rares. En particulier, il est fréquent de considérer que la productivité des salariés stagne, voire décroît, à partir de 50 ans ce qui induirait une moindre demande des firmes pour ces salariés, la baisse de productivité n'étant pas compensée par une baisse du coût salarial du fait de rigidités.

Cette hypothèse généralement admise de diminution de la productivité des salariés âgés n'est cependant fondée que sur des témoignages de managers ou dirigeants d'entreprises, ou sur les

¹ INSEE - Département des études économiques d'ensemble - Division « Marchés et Stratégies d'entreprise »
15, Boulevard Gabriel Péri - BP 100 - 92244 Malakoff Cedex - Mail : patrick.aubert@insee.fr

résultats de tests psychométriques dans le domaine de la psychologie. L'estimation empirique du lien entre âge et productivité est en revanche beaucoup plus rare, principalement du fait de la difficulté de trouver des données qui fournissent une mesure satisfaisante de la productivité et permettent cette estimation. Dans les estimations les plus anciennes, telle celle de Medoff et Abraham (1981), la mesure de la productivité utilisée est généralement une évaluation de la performance des salariés effectuée par leurs supérieurs ou leurs collègues. La pertinence de ces évaluations comme mesure de la productivité n'est cependant pas évidente, et les observations sont généralement conduites que sur un nombre très restreint d'entreprises (deux dans le cas de Medoff et Abraham), de telle sorte qu'il se pose un problème de généralisation des résultats important.

A l'exception de telles études, difficiles à généraliser et dont la manière d'appréhender la productivité peut être mise en cause, l'estimation d'une relation entre âge et productivité n'est en fait l'objet que d'une littérature très récente. Dans cette littérature, la productivité n'est plus évaluée au niveau individuel mais au niveau de catégories de salariés définies par certaines caractéristiques semblables. La productivité d'un groupe de salarié est alors définie comme la « contribution » de ce groupe à la valeur ajoutée de l'entreprise, et les estimations sont conduites sur des bases appariées employeur-employé contenant des données sur les caractéristiques financières de l'entreprise et la structure de leur main d'œuvre. La première étude adoptant cette méthodologie est celle de Hellerstein, Neumark et Troske (1999). A partir d'un panel contenant plus de 3000 firmes et 120 000 salariés, ils estiment la productivité marginale relative de groupes de salariés définis par le sexe, l'âge, la « race » et la qualification. Afin de tester l'égalité entre salaire et productivité marginale, ces auteurs estiment également des équations de salaire. Leur résultat, pour ce qui concerne le lien entre âge, salaire et productivité, est que le salaire plus élevé des travailleurs plus âgés est bien le reflet d'une productivité supérieure.

Deux études ultérieures viennent néanmoins nuancer ce résultat. Haegeland et Klette (1999), sur données norvégiennes, et Crépon, Deniau, Perez-Duarte (2001), sur données françaises, n'estiment pas en effet une productivité supérieure pour les travailleurs âgés, comme c'était le cas dans Hellerstein, Neumark et Troske. Crépon, Deniau et Perez-Duarte proposent de surcroît une nouvelle méthode pour tester directement l'égalité entre salaire et productivité, en estimant le « rabais relatif » de chaque catégorie de travailleur. De leurs estimations, effectuée sur un panel réunissant 75 000 entreprises pérennes entre 1994 et 1997, il ressort que les travailleurs de plus de 50 ans seraient plus payés relativement aux plus jeunes que leur différence de productivité laisserait supposer, et cela aussi bien dans l'industrie que dans les services. Ces résultats sont de plus confirmés par certains tests de robustesse à travers le temps, le secteur d'activité et les hypothèses retenues.

Ces estimations peuvent néanmoins souffrir de deux types de problème. Il y a tout d'abord un problème de précision, dû à la qualité des données. Par exemple, la part du travail représentée par chaque catégorie de travailleurs dans Hellerstein, Neumark, Troske est mesurée par la part de cette catégorie dans les effectifs, estimée sur un échantillon de ces derniers. Ce problème n'est cependant pas présent dans Crépon, Deniau, Perez-Duarte puisque les données utilisées sont exhaustives.

Un second problème vient de la difficulté d'estimation des fonctions de productions. De telles estimations sont susceptibles d'être fortement biaisées du fait de la simultanéité des variations de la

production et des ajustements des facteurs de production, comme cela est documenté dans Griliches et Mairesse (1997). En particulier, l'estimation de la productivité marginale relative des différentes catégories d'âge peut souffrir d'un tel biais, si les politiques d'embauche et de licenciement des entreprises ne sont pas identiques pour chaque classe d'âge et que ces entreprises ajustent leur emploi en anticipant la variation de production. Une entreprise en mauvaise santé économique aura par exemple tendance à ne pas embaucher pendant plusieurs années consécutives, laissant vieillir sa main-d'œuvre en place, si bien que l'on observera de manière conjointe une production en baisse et une augmentation du poids des classes d'âge élevé. Au contraire, si une entreprise prévoit des chocs positifs sur sa production, elle embauchera un certain nombre de travailleurs jeunes, et on observera dans ce cas une production en hausse et une diminution de la part des plus âgés dans l'effectif total. En d'autres termes, un problème se pose quant au sens de la causalité : les entreprises sont-elles moins productives parce qu'elles emploient plus de travailleurs âgés, ou au contraire emploient-elles plus de travailleurs âgés parce qu'elles sont moins productives ?

La prise en compte de ce problème d'endogénéité est l'objet principal de cette étude. Nous adoptons pour cela une démarche similaire à celle de Crépon, Deniau, Perez-Duarte, à la fois pour ce qui est de la méthode et des données. Nous estimons des équations de salaire et de production similaires sur un panel d'entreprises pérennes sur la période 1994-2000, construit à partir d'un appariement BRN-DADS. En revanche, afin de sortir de l'opposition entre salariés « âgés » ou « jeunes » et une masse de salariés d'âge moyen, nous distinguons huit classes d'âge différentes définies par tranches de cinq ans et tentons ainsi d'estimer une « profil » âge-productivité plus complet.

Pour anticiper sur la présentation des résultats, les diverses estimations que nous avons conduites soulignent l'importance du problème d'endogénéité. Différentes hypothèses sur la forme du terme individuel inobservable dans la fonction de production conduisent à des résultats différents. En particulier, les résultats des diverses études différentes, aussi bien Hellerstein, Neumark et Troske (1999) que Crépon, Deniau, Perez-Duarte (2001) sont susceptibles d'être affectés par ce biais. Après avoir discuté le choix de la méthode d'Arellano et Bond (1991), nous estimons un profil de la productivité croissant et concave avec l'âge, la croissance de la productivité diminuant à partir de 40 ans pour s'annuler après 50 ans. Ce résultat est robuste à diverses spécifications et nous conduit à rejeter l'idée que la moindre demande de travailleurs âgés par les entreprises résulte d'une productivité plus faible de ces travailleurs.

L'organisation du document est la suivante : la section 2 présente le modèle et une discussion du biais d'endogénéité. Les données sont présentées dans la section 3. Les résultats sont développés dans les sections 4 et 5, et la conclusion dans la section 6.

2. Le modèle

Fonctions de production

Le cadre théorique utilisé ici est similaire à celui de Crépon et al (2001), lui-même inspiré de la méthode initialement utilisée par Hellerstein et al. Considérons une fonction de production de type Cobb-Douglas :

$$\ln(Q_i) = \ln(A) + \beta \cdot \ln(K_i) + \alpha \cdot \ln(\lambda \cdot L_i) + \varepsilon_i$$

où Q_i désigne la valeur ajoutée dans la firme i , K_i son capital et $\lambda \cdot L$ le « travail efficace ».

L'objectif étant l'estimation des productivités relatives de divers sous-groupes de travailleurs, $\lambda \cdot L$ peut être réécrit comme la somme du travail efficace relatif de chacun de ces sous-groupes, à l'exception d'un sous-groupe choisi comme référence :

$$\lambda L = \sum_i \lambda_i \cdot L_i = \lambda_0 \cdot L \left(\frac{L_0}{L} + \sum_{i \neq 0} \frac{\lambda_i}{\lambda_0} \cdot \frac{L_i}{L} \right) = \lambda_0 \cdot L \cdot \left(1 + \sum_{i \neq 0} \left(\frac{\lambda_i}{\lambda_0} - 1 \right) \cdot \frac{L_i}{L} \right)$$

La fonction de production se réécrit alors

$$\ln(Q) = \text{cons} + \beta \cdot \ln(K) + \alpha \cdot \log(L) + \alpha \cdot \log \left(1 + \sum_{j \neq 0} \left(\frac{\lambda_j}{\lambda_0} - 1 \right) \cdot \frac{L_j}{L} \right) + \varepsilon.$$

et si, à l'instar de Crépon et al, on retient l'hypothèse que la productivité relative d'un sous-groupe défini par plusieurs caractéristiques démographiques et égal à la somme des productivités relatives des sous-groupes correspondants à chacune de ces caractéristiques prises séparément (par exemple, si la productivité relative du groupe des femmes âgées non-qualifiées peut se réécrire comme la somme d'un terme « productivité relative des femmes », d'un terme « productivité relative des âgés » et d'un terme « productivité relative des non-qualifiés »

$$\frac{\lambda_{f\hat{e}, \hat{a}g\acute{e}, non-qualif}}{\lambda_0} = \frac{\lambda_{fe}}{\lambda_0} + \frac{\lambda_{\hat{a}g\acute{e}}}{\lambda_0} + \frac{\lambda_{non-qualif}}{\lambda_0}$$

alors la fonction de production se linéarise simplement :

$$\ln(Q_i) = \text{cons} + \beta \cdot \ln(K_i) + \alpha \cdot \log(L_i) + \sum_{j \neq 0} \alpha \cdot \left(\frac{\lambda_j}{\lambda_0} - 1 \right) \frac{L_j}{L_i} + \varepsilon_i$$

Il s'agit d'une fonction de production classique dans laquelle on a ajouté parmi les régresseurs la part du travail total effectuée par chaque sous-groupe de salariés considéré, à l'exception de l'un d'entre eux choisi comme référence. Hellerstein et al (1999) et Crépon et al (2001) présentent leurs résultats à partir de l'estimation en « between » de ces fonctions, c'est à dire une estimation à partir de la moyenne des variables sur plusieurs années. En d'autres termes, un groupe de salarié sera estimé plus productif que le groupe de référence si une entreprise produit plus en moyenne qu'une entreprise comparable (en terme de niveau de capital et de travail, de taille, secteur, âge) lorsque ce groupe représente une part plus grande de ses effectifs.

Crépon et al. estiment également les versions en « première différence » et « différence longue » de ces fonctions, ainsi qu'un modèle en niveaux instrumentés par les différences passées des variables. L'interprétation d'une fonction de production estimée en différence est un peu différente de celle d'une fonction de production estimée en niveau, comme c'était le cas dans le cas « between ». Dans un modèle en différence, un groupe de salariés est estimé plus productif si, en moyenne, la production augmente plus dans une entreprise que dans une firme comparable lorsque la part du groupe dans l'emploi augmente plus dans la première entreprise que dans la seconde.

Le problème d'endogénéité

Néanmoins, comme nous l'avons exprimé en introduction, l'estimation de fonctions de productions est sujette à un certain nombre de biais, du fait en particulier de la simultanéité des décisions d'embauche et d'ajustement du capital avec des termes non-observables qui influent sur la production.

Considérons par exemple que le résidu pour l'entreprise i dans l'équation ci-dessus peut s'écrire $\varepsilon_{i,t} = u_{i,t} + v_{i,t}$, où $u_{i,t}$ représente le niveau de productivité générale de l'entreprise i à l'instant t , tandis que $v_{i,t}$ est un choc transitoire sur la production. $u_{i,t}$ peut suivre une marche aléatoire $u_{i,t} = u_{i,t-1} + \eta_{i,t}$, $\eta_{i,t}$ représentant un choc durable à l'instant t sur la productivité générale des facteurs de l'entreprise i . Si $\eta_{i,t}$ est anticipé à l'instant $(t-1)$, l'entreprise ajuste sa main-d'œuvre et son capital en conséquence, si bien que $\log L_{i,t} = \log L_{i,t-1} + \eta_{i,t}$ et $\log K_{i,t}$ seront certainement corrélés avec $\eta_{i,t}$. Sous cette hypothèse, le terme individuel inobservable $u_{i,t} = \eta_{i,t}$ sera corrélé non seulement avec $\log L_{i,t}$ et $\log K_{i,t}$ mais également avec tous leurs niveaux et différences passées.

Si le résidu de l'équation de production s'écrit bien sous cette forme, c'est à dire s'il contient un terme idiosyncrasique suivant une marche aléatoire dont l'innovation est anticipée par l'entreprise et donc corrélée aux ajustements de main d'œuvre et de capital, alors non seulement l'estimation en niveau, mais également celle des modèles en « première différence » ou « différence longue », seront biaisés. En effet, la fonction de production ci-dessus en différence s'écrit :

$$\ln(Q)_{i,t} = \beta \cdot \ln(K)_{i,t} + \alpha \cdot \log(L)_{i,t} + \sum_{j=0}^{\lambda_j} \alpha \cdot \left(\frac{\lambda_j}{\lambda_0} - 1 \right) \frac{L_j}{L}_{i,t} + \eta_{i,t} + v_{i,t} - v_{i,t-1}$$

et les variables $\ln(K)_{i,t}$ et $\log(L)_{i,t}$ sont toujours corrélées au terme inobservable $\eta_{i,t}$. Notons que l'utilisation des GMM selon la procédure d'Arellano et Bover (1995) c'est à dire en instrumentant les niveaux des variables par les différences passées, ne suffisent pas non plus pour se défaire du biais de simultanéité, puisque l'équation se réécrit

$$\ln(Q)_{i,t} = \beta \cdot \ln(K)_{i,t} + \alpha \cdot \log(L)_{i,t} + \sum_{j \neq 0} \alpha \cdot \left(\frac{\lambda_j}{\lambda_0} - 1 \right) \frac{L_j}{L_0} + \eta_{i,s} + v_{i,t}$$

et que les niveaux ou les différences passés de l'emploi et du capital seront également corrélés aux innovations passées $\eta_{i,s}$ ($s < t$).

Le même problème se pose par ailleurs en ce qui concerne les proportions de chacune des classes d'âge dans l'emploi total. La variation de ces proportions $PL_{i,t}^j = \frac{L_{i,t}^j}{L_{i,t}} - \frac{L_{i,t-1}^j}{L_{i,t-1}}$ sera en effet corrélée avec le terme $\eta_{i,t}$, représentant un « choc durable anticipé » sur la productivité, non seulement par le biais de la variation de son dénominateur $L_{i,t}$, mais également parce que les embauches et licenciements par classe d'âge peuvent être directement corrélés avec ce terme. C'est le cas par exemple si, face à un choc positif sur sa demande, l'entreprise augmente sa main d'œuvre en embauchant de nombreux jeunes.

Le tableau suivant donne une idée de ce biais en présentant les corrélations entre variation de l'emploi et du capital d'une part, et variation des parts des différentes catégories d'autre part.

Coefficient de corrélation <i>p-value</i>	Variation entre (t-1) et t									
	<25 ans	25 à 29	30 à 34	40 à 44	40 à 44	50 à 54	>55 ans	Q0	Non-qualifiés	Très qualifiés
LOGL										
Variation entre (t-1) et t	0,071	0,021	-0,006	-0,017	-0,028	-0,024	0,000	0,028	0,029	-0,043
	<,001	<,001	0,008	<,001	<,001	<,001	0,869	<,001	<,001	<,001
LOGK										
Variation entre (t-1) et t	0,014	0,013	0,004	-0,003	-0,003	-0,013	-0,011	-0,002	0,010	-0,020
	<,001	<,001	0,087	0,249	0,136	<,001	<,001	0,288	<,001	<,001

Champ : 33 896 entreprises de l'industrie observées entre 1996 et 2000

A l'exception de la classe d'âge des plus de 55 ans, la variation de la part de chacune des catégories d'âge et de qualification est corrélée de manière significative avec la variation du niveau d'emploi. De plus, cette corrélation est positive pour les catégories des plus jeunes et des moins qualifiés, et négative pour les plus âgés et les plus qualifiés. Si l'on accepte l'idée d'une causalité inverse entre production et niveau des inputs, c'est à dire si l'on accepte l'idée que les entreprises peuvent anticiper un choc de demande positif ou négatif et ajuster simultanément leur niveau d'emploi en embauchant ou licenciant, alors les coefficients associés aux variations des parts des différentes catégories d'âge ou de qualification seront biaisés dans l'estimation de la fonction de production. Ce biais sera de plus positif pour les catégories plus jeunes et moins qualifiées, et négatif pour les plus anciens et plus qualifiés.

Ce problème de biais peut être résolu par l'utilisation de la méthode des moments généralisés (GMM) selon la procédure d'Arellano et Bond (1991), dans laquelle l'équation est estimée en

différence première en utilisant comme instruments les niveaux passés. L'hypothèse sous-jacente est que ces variables passées ne sont pas corrélées à l'innovation présente $\eta_{i,t}$ du terme inobservable $u_{i,t}$. En d'autres termes, nous estimons la fonction de production en différence suivante

$$\ln(Q)_{i,t} = \beta \cdot \ln(K)_{i,t} + \alpha \cdot \log(L)_{i,t} + \sum_{j \neq 0} \gamma_j \cdot \frac{L_j}{L}_{i,t} + \mu_{i,t}$$

avec $\mu_{i,t} = \eta_{i,t} + v_{i,t} - v_{i,t-1}$

avec les conditions d'orthogonalité

$$E \left(\ln(K)_{i,t-s} \cdot (\eta_{i,t} + v_{i,t} - v_{i,t-1}) \right) = 0$$

$$E \left(\log(L)_{i,t-s} \cdot (\eta_{i,t} + v_{i,t} - v_{i,t-1}) \right) = 0$$

$$E \left(\frac{L_j}{L}_{i,t-s} \cdot (\eta_{i,t} + v_{i,t} - v_{i,t-1}) \right) = 0$$

pour toute entreprise i à toute date t d'observation et pour toutes les catégories de salariés j considérées. Les retards s considérés sont par ailleurs supérieurs à deux. Nous ajoutons également des indicatrices de secteur, d'année, de taille d'établissement.

Cette méthode donne généralement des résultats peu satisfaisants, du fait notamment de la mauvaise qualité des instruments utilisés. La situation est cependant particulière dans notre cas, du moins en ce qui concerne l'estimation des coefficients associés aux variations de la proportion de diverses classes d'âge dans l'emploi total de la firme. Ces variations ne résultent pas en effet uniquement des ajustements de la main d'œuvre. Le vieillissement des individus fournit en fait une source de variabilité importante et qui n'est pas a priori corrélée à des variations anticipées de la productivité générale de l'entreprise. La structure par âge passée de l'entreprise constituera donc un « bon » instrument de la variation de la taille relative des classes d'âge. A titre d'exemple, le tableau ci-dessous présente le R2 de la régression de chacune des variables explicatives de la fonction de production sur l'ensemble de ces variables observées deux et trois années auparavant. Pour la plupart des variables, la corrélation avec les instruments est faible, ce qui est un résultat classique dans l'estimation des fonctions de production. On constate néanmoins que les instruments ont un bon pouvoir prédictif pour ce qui concerne la variation de la part des classes d'âge.

Variable	R2
DLOGL	0,011
DLOGK	0,022
DPLA1	0,082
DPLA2	0,126
DPLA3	0,162
DPLA5	0,176
DPLA6	0,181

DPLA7	0,173
DPLA8	0,143
DPLQ1	0,023
DPLQ3	0,013
Régession par les moindres carrés ordinaires de la variation entre 1998 et 1999 de chacune des variables considérées sur LOGL, LOGK et les parts de chacune des catégories d'âge et de qualification en 1996 et en 1997. Champ = 33 896 entreprises de l'industrie	

Une spécification plus générale pour la fonction de production

La forme Cobb-Douglas retenue pour la fonction de production est assez restrictive. Elle ne permet pas en particulier de prendre en compte les complémentarités ou substituabilités entre les différents facteurs de production. Par exemple, elle implique que la complémentarité entre capital et travail soit la même pour tous les types de travail. Cette spécification trop restrictive est susceptible de biaiser les coefficients estimés pour les différents facteurs. Il est possible de la rendre moins flexible en estimant également les coefficients du second degré d'une fonction de type translog, c'est à dire en ajoutant parmi les régresseurs les produits des différents facteurs.

Tester l'égalité du salaire et de la productivité marginale

En appliquant une démarche similaire à celle du cas des fonctions de production, où les productivités relatives des catégories de salaires étaient estimées via la relation entre la part de cette catégorie dans l'emploi et la production totale, les salaires relatifs de chaque catégorie peuvent également être estimés via la relation entre part dans l'emploi et salaire moyen dans l'entreprise. Le coût salarial agrégé dans l'entreprise s'écrivant comme la somme des coûts salariaux pour chacune des catégories, le coût horaire moyen est

$$\bar{w} = \frac{\sum_i w_i \cdot L_i}{L_i} = w_0 \cdot \left(1 + \sum_{i \neq 0} \left(\frac{w_i}{w_0} - 1 \right) \frac{L_i}{L} \right)$$

Sous l'hypothèse de coûts salariaux relatifs des catégories constant d'une entreprise à l'autre, ces coûts relatifs peuvent être estimés par l'équation de salaire suivante

$$\ln(\bar{w}_i) = \text{cons} + \ln \left(1 + \sum_{j \neq 0} \left(\frac{w_j}{w_0} - 1 \right) \frac{L_j}{L} \right) + v_i$$

ou par la forme en différence première linéarisée suivante

$$\ln(\bar{w}_i) = \sum_{j \neq 0} \left(\frac{w_j}{w_0} - 1 \right) \frac{L_j}{L} + \varepsilon_i$$

L'estimation conjointe d'une fonction de production et d'une fonction de salaire était l'idée principale dans Hellerstein, Neumark, Troske. Cette estimation doit permettre de tester ensuite l'adéquation entre salaire et productivité, en comparant les salaires relatifs $\frac{w_j}{w_0}$ et productivités relatives $\frac{\lambda_j}{\lambda_0}$ des différentes catégories de travailleurs retenues.

Cette méthode peut néanmoins poser problème, notamment si l'une des deux équations est biaisée ou si les deux équations subissent des biais de sens ou d'ampleur différente. En particulier, la comparaison des salaires relatifs et productivités relatives impose que ces grandeurs soient exprimées comme des ratios du salaire ou de la productivité de référence. Puisque nous estimons une forme linéaire de la fonction de production, ce ratio doit être recalculé à partir des paramètres estimés dans le cas de la productivité relative

$$\frac{\lambda_i}{\lambda_0} = 1 + \frac{\hat{\gamma}_i}{\hat{\alpha}}$$

Où $\hat{\alpha}$ est le coefficient estimé pour le nombre total d'heures travaillées et $\hat{\gamma}_i$ le coefficient estimé pour la part de la catégorie considérée dans l'emploi. Si $\hat{\alpha}$ est biaisé, tous les ratios salaire sur productivité reconstruits seront faussés.

Les données disponibles dans l'appariement DADS-BRN étant plus riches que celles dont disposent Hellerstein, Neumark, Troske, Crépon et al. imaginent une manière plus directe de tester l'adéquation entre salaire et productivité. Leur méthode se fonde sur l'idée de faire apparaître directement dans la fonction de production la part dans les coûts salariaux, et non plus dans le travail, des diverses catégories de salariés. S'il y a adéquation entre salaire et productivité, il ne devrait normalement pas y avoir de corrélation entre production et part dans les coûts salariaux puisque toute augmentation de la part d'une catégorie d'âge dans les salaires traduit une augmentation de la productivité de même ampleur.

Le « travail efficace » dans la fonction de production se réécrit ainsi

$$\begin{aligned} \lambda L &= \lambda_i \cdot L_i = \frac{\lambda_0}{w_0} \cdot (wL) \frac{w_0 L_0}{wL} + \frac{\lambda_i/w_i}{\lambda_0/w_0} \cdot \frac{w_i L_i}{wL} \\ &= \frac{\lambda_0}{w_0} \cdot (wL) \left(1 + \left(\frac{\lambda_i/\lambda_0}{w_i/w_0} - 1 \right) \cdot \frac{w_i L_i}{wL} \right) \end{aligned}$$

ce qui conduit à la forme linéarisée de la fonction de production

$$\ln(Q_i) = \text{cons} + \beta \cdot \ln(K_i) + \alpha \cdot \log((wL)_i) + \alpha \cdot \left(\frac{\lambda_j/\lambda_0}{w_j/w_0} - 1 \right) \frac{(wL)_j}{wL}_i + \varepsilon_i$$

et l'on estime bien directement le rapport de la productivité relative du groupe j sur son salaire relatif.

3. Les données

Nos données initiales sont un appariement des Déclarations Administratives de Données Sociales (DADS) et des Bénéfices Réels Normaux (BRN) de 1994 à 2000. La valeur ajoutée, utilisée comme variable dépendante dans nos régressions, est mesurée comme la production moins les consommations intermédiaires et les variations de stock dans les BRN. Le capital est mesuré par les immobilisations productives brutes. Les DADS fournissent notre mesure du travail, défini comme le nombre total d'heures travaillées dans l'entreprise au cours de l'année. Le coût salarial total est calculé de même à partir des DADS. Pour chaque entreprise, on calcule également ces grandeurs (travail et coût salarial agrégés) pour des « cellules » définies par des catégories de qualification² (4 catégories) et d'âge (8 catégories, par tranche de cinq ans).

Un certain nombre de filtrages ont par ailleurs été effectués, correspondant aux conditions suivantes : présence de l'entreprise dans chacune des bases pour chaque année de 1994 à 2000 ; l'effectif, la valeur ajoutée, les immobilisations productives brutes doivent être strictement positives pour chacune des années (dans les BRN) ; le coût salarial total doit être strictement positif dans les DADS et présent dans la base BRN. Le champ est celui des entreprises individuelles, grandes entreprises nationales, associations ou sociétés. Les secteurs « pêche » et « administrations » sont exclus. Afin d'exclure les valeurs aberrantes, nous éliminons également de la base les entreprises qui se trouvent dans le premier ou le dernier centile de la distribution pour le coût moyen horaire, le nombre moyen d'heures travaillées par jour, la moyenne de 1994 à 2000 de la productivité du travail et de celle du capital, ainsi que pour la croissance de l'effectif, du capital ou de la production.

A l'issue de ces divers nettoyages, la base compte 33 896 entreprises pour l'industrie, 22 387 dans la construction, 68 780 dans le commerce et 51 861 dans les services.

Le tableau suivant présente quelques statistiques descriptives pour chacun de ces secteurs. Une distribution par tranche de taille d'effectif et par année de création est par ailleurs présentée en annexe.

² Nous adoptons un découpage par niveaux de qualification simple et distinguons trois grandes catégories de salariés, définies à partir de la catégorie socioprofessionnelle. La catégorie des « très qualifiés » regroupe les ingénieurs, techniciens, managers et professions intermédiaires. Les « qualifiés » sont les ouvriers et employés qualifiés. Enfin, les « non qualifiés » regroupent les ouvriers non-qualifiés, ainsi que les agents de surveillance, employés de commerce et personnels des services directs aux particuliers.

Statistiques descriptives sur les diverses catégories de travailleurs

Secteur	Industrie			Construction			Commerce			Services		
	Moy.	Std	Median	Moy.	Std	Median	Moy.	Std	Median	Moy.	Std	Median
Capital (log)	8,055	1,728	7,844	6,778	1,335	6,724	7,170	1,372	7,159	7,172	1,590	7,107
Nombre d'heures travaillées (log)	10,368	1,338	10,235	9,838	1,068	9,761	9,401	1,108	9,256	9,667	1,258	9,535
<25	0,098	0,122	0,060	0,115	0,130	0,079	0,126	0,159	0,067	0,120	0,162	0,056
25-29	0,144	0,131	0,125	0,127	0,134	0,106	0,153	0,172	0,118	0,171	0,171	0,143
30-34	0,159	0,131	0,147	0,149	0,144	0,131	0,152	0,172	0,123	0,164	0,166	0,141
40-44	0,140	0,127	0,127	0,138	0,142	0,119	0,132	0,170	0,086	0,125	0,157	0,089
45-49	0,135	0,131	0,118	0,137	0,144	0,116	0,125	0,172	0,067	0,119	0,163	0,071
50-54	0,102	0,120	0,077	0,109	0,137	0,078	0,097	0,160	0,010	0,091	0,151	0,023
>55	0,073	0,107	0,041	0,080	0,119	0,031	0,074	0,144	0,000	0,069	0,135	0,000
Stagiaires et apprentis	0,023	0,065	0,000	0,039	0,080	0,000	0,045	0,105	0,000	0,027	0,083	0,000
Non qualifiés	0,256	0,240	0,198	0,121	0,174	0,045	0,279	0,286	0,187	0,212	0,296	0,029
Très qualifiés	0,260	0,209	0,213	0,194	0,191	0,153	0,365	0,281	0,313	0,383	0,321	0,314
Femmes	0,270	0,234	0,202	0,086	0,103	0,065	0,446	0,326	0,367	0,432	0,309	0,413
<i>Observations</i>												
Nb d'entreprises	33 896			22 387			68 780			51 861		
Nb d'années	7			7			7			7		

4. Résultats

Tous les résultats présentés ci-dessous ont été estimés sur les données des entreprises de l'industrie. Les secteurs du commerce, de la construction et des services mènent à des résultats similaires et les profils âge-productivité estimés pour ces secteurs sont présentés en annexe.

Méthode d'estimation et biais d'endogénéité

La principale difficulté dans l'estimation de fonctions de production réside dans la corrélation entre les variables, c'est à dire les facteurs de production de l'entreprise, et un terme individuel, inobservable pour l'économètre mais dont les évolutions peuvent être anticipées par l'entreprise. La forme que peut prendre ce terme inobservable, ainsi que le biais qui en résulte, ont été discutés dans la partie précédente. En particulier, nous avons cherché à justifier l'idée que l'estimation par les MCO ou l'estimation d'un modèle en niveau est susceptible de conduire à des coefficients trop élevés pour les catégories jeunes et non-qualifiées, et trop bas pour les âgés et les plus qualifiés.

Afin d'illustrer ce propos, nous présentons dans le tableau suivant le profil âge-productivité résultant de l'estimation des équations en niveau

$$\log(Q)_{i,t} = \beta \cdot \log(K)_{i,t} + \alpha \cdot \log(L)_{i,t} + \sum_{j \neq 0} \gamma_j \cdot \frac{L_j}{L}_{i,t} + \vartheta_{i,t}$$

et en différence

$$\log(Q)_{i,t} = \beta \cdot \log(K)_{i,t} + \alpha \cdot \log(L)_{i,t} + \sum_{j=0}^{\gamma} \gamma_j \cdot \frac{L_j}{L_{i,t}} + \mu_{i,t}$$

Nous estimons de manière jointe ces équations pour chacune des années d'observation par la méthode des moments généralisées (MMG), afin de tenir compte d'une éventuelle hétéroscédasticité du fait de la corrélation entre les résidus d'une année sur l'autre. A chaque fois, nous considérons d'abord que les facteurs de production sont exogènes et les utilisons donc comme leurs propres instruments, puis nous réestimons l'équation avec les valeurs passées des variables comme instruments. Nous utilisons les niveaux passés des variables comme instruments dans le cas de l'équation en différence, selon la méthodologie proposée par Arellano et Bond (??), et les différences passées dans l'équation en niveau, selon la méthode d'Arellano et Bover (??).

	param	stde	param	stde	param	stde	param	stde
Capital	0,207	0,006	0,163	0,006	0,268	0,018	0,089	0,026
Nombre d'heures travaillées	0,791	0,014	0,221	0,008	0,769	0,019	0,324	0,057
Age								
<25 ans	-0,058	0,067	0,011	0,011	0,067	0,045	-0,569	0,067
25-29	0,107	0,064	0,029	0,010	0,107	0,028	-0,415	0,053
30-34	0,119	0,069	0,009	0,008	0,085	0,025	-0,164	0,029
35-39	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
40-44	0,047	0,069	-0,018	0,008	-0,018	0,026	0,132	0,025
45-49	0,026	0,065	-0,058	0,011	0,015	0,026	0,225	0,032
50-54	0,005	0,065	-0,060	0,013	0,072	0,029	0,283	0,040
>55 ans	0,006	0,072	-0,051	0,015	0,144	0,039	0,278	0,044
Niveau de qualification								
Stagiaires, apprentis	-0,649	0,035	-0,174	0,012	-0,629	0,060	-0,220	0,065
Non-qualifiés	-0,144	0,012	0,015	0,005	0,025	0,029	0,087	0,031
Très qualifiés	0,353	0,024	-0,015	0,008	-0,040	0,041	-0,336	0,063
Sexe								
Femme	-0,022	0,015	-0,007	0,016	0,089	0,045	0,092	0,088

Variables	X_t	$X_t - X_{t-1}$	X_t	$X_t - X_{t-1}$
Instruments	$Z_t = X_t$	$Z_t = X_t - X_{t-1}$	$Z_t = X_{t-1} - X_{t-2}$	$Z_t = X_{t-2}$
Statistique de Sargan	139,2	98,6	190,0	109,7
Degrés de liberté	52	52	52	52
P-valeur	0	0	0	0

Champ : 33 896 entreprises pérennes du secteur de l'industrie observées de 1996 à 2000. Dans une première étape, les moyennes pour chaque année et dans chacun des secteurs (selon la classification NES114) des variables explicatives et de la variable dépendante ont été soustraites à ces variables .

La première observation concerne la spécification du modèle. Toutes les spécifications sont rejetées par le test de sur-identification de Sargan, qui teste de manière jointe l'orthogonalité des instruments avec les résidus pour chacune des années d'observation. En d'autres termes, il s'agit d'un test de l'exogénéité des instruments. Dans la mesure où nous estimons le modèle de manière jointe

pour chacune des années et en imposant que les paramètres soient les mêmes d'une année sur l'autre, le test de Sargan peut également être interprété comme un test de cohérence temporelle du modèle. La valeur élevée de la statistique de Sargan sous toutes les spécifications pourrait s'interpréter comme un rejet de l'hypothèse que les productivités marginales des différentes variables sont constantes dans le temps.

Les productivités marginales du travail et du capital conduisent à des rendements d'échelles à peu près constants dans les estimations des modèles en niveau, et nettement décroissants dans les estimations des modèles en différences. Cette « atténuation » des coefficients du capital et du travail est un résultat classique dans l'estimation de fonctions de production en différence. Par ailleurs, la part de femmes dans l'emploi ne semble pas avoir d'impact significatif sur la production et les coefficients des parts des catégories de qualification signifieraient, à l'exception du cas du modèle en niveau sous l'hypothèse d'exogénéité des variables, que la productivité marginale relative est d'autant plus faible que la qualification est élevée. Ce résultat assez étonnant et difficilement crédible correspond certes bien à l'idée développée précédemment d'un biais positif sur la productivité marginale estimée pour les non-qualifiés et négatif pour celle des très qualifiés, mais il reste observé dans le modèle estimé selon la méthode d'Arellano et Bond, censée corriger ce biais.

En revanche, les résultats semblent cohérents avec notre hypothèse d'un biais sur l'estimation des productivités marginales par âge, positif pour les plus jeunes et négatif pour les plus âgés. Dans les trois premiers modèles, la productivité est maximale pour les classes d'âge entre 20 et 30 ans. Dans les deux modèles qui acceptent l'exogénéité des variables explicatives, elle est même strictement décroissante avec l'âge jusqu'à 50 ans (si l'on excepte la classe d'âge de référence pour le modèle en niveau, qui semble plutôt indiquer que les coefficients estimés pour toutes les autres classes d'âges subissent un biais positif), puis constante au-delà. Dans le modèle en niveau instrumenté par les différences passées, la productivité décroît avec l'âge jusque vers 40 ans, puis croît assez fortement au-delà de 45 ans. Au contraire, dans le cas du modèle en différences premières, qu'on suppose tenir compte du biais d'endogénéité, les coefficients estimés tracent un profil âge-productivité nettement croissant et concave. La productivité ne croîtrait pas après 50 ans, mais elle ne diminuerait pas non plus.

Le test de Sargan ne nous permet pas de justifier l'utilisation d'un modèle plutôt qu'un autre, puisque l'exogénéité des instruments sur chacune des années d'observation ne peut être acceptée pour aucune des spécifications. Nous ne nous appuyons donc que sur la discussion théorique de la partie précédente et sur le caractère robuste des estimations conduites avec la méthode d'Arellano et Bond pour retenir ces estimations dans la suite. Le tableau suivant illustre ce caractère en présentant de telles estimations sous plusieurs spécifications.

	param	stde	param	stde	param	stde	param	stde	param	stde
Variable dépendante	DLOGQ		DLOGQ		DLOGQ		DLOGQ		DLOGQ	
Capital	0,215	0,018	0,210	0,018	0,239	0,016	0,277	0,017	0,214	0,021
Nb d'heures travaillées	0,424	0,030	0,512	0,029	0,543	0,025	0,520	0,023	0,502	0,027
<i>Age</i>										

<25 ans	-0,438	0,036	-0,366	0,036	-0,327	0,032	-0,263	0,032	-0,371	0,036
25-29	-0,287	0,028	-0,237	0,028	-0,205	0,024	-0,193	0,026	-0,245	0,027
30-34	-0,114	0,018	-0,086	0,017	-0,072	0,015	-0,073	0,016	-0,090	0,017
35-39	0	0	0	0	0	0	0	0	0	0
40-44	0,039	0,022	0,069	0,021	0,041	0,020	0,066	0,016	0,077	0,019
45-49	0,090	0,025	0,106	0,025	0,061	0,023	0,096	0,022	0,116	0,023
50-54	0,123	0,029	0,131	0,030	0,079	0,026	0,119	0,028	0,145	0,029
>55 ans	0,109	0,031	0,127	0,032	0,075	0,029	0,128	0,031	0,138	0,032
Niveau de qualification										
Stagiaires, apprentis			-0,425	0,054	-0,467	0,050	-0,384	0,056	-0,398	0,054
Non-qualifiés			0,051	0,027	0,037	0,026	0,124	0,026	0,028	0,131
Très qualifiés			-0,233	0,050	-0,21	0,047	-0,189	0,050	0,123	0,276
Sexe										
Femme			-0,045	0,072	-0,046	0,066	0,339	0,075	-0,022	0,071
Termes croisés										
P(non-qual)*LOGK									0,006	0,016
P(très-qual)*LOGK									-0,056	0,043
<hr/>										
Variables	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$	$X_t - X_{t-1}$
Instruments	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2}, X_{t-3}, X_{t-4}, \text{ et } X_{t-5}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$	$Z_t = X_{t-2} \text{ et } X_{t-3}$
Centrage par secteur	oui	oui	oui	non	non	non	non	non	oui	oui
<hr/>										
Statistique de Sargan	180,9	257,3	357,8	390,9	390,9	390,9	390,9	390,9	294,8	294,8
Degrés de liberté	72	104	169	129	129	129	129	129	120	120
P-valeur	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
<hr/>										
Nombre d'entreprises	33896	33896	33896	33896	33896	33896	33896	33896	33896	33896
Début des observations	1996	1996	1996	1996	1996	1996	1996	1996	1996	1996
Fin des observations	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Estimation par la méthode des moments généralisés (deuxième étape, matrice optimale). La quantité de travail est mesurée comme le nombre total d'heures travaillées dans l'entreprise et dans l'année. Les variables endogènes sont les variations entre (t-1) et (t) des grandeurs considérées. Les instruments sont les niveaux retardés en (t-2) et (t-3) de ces grandeurs. La variation du log des indices de prix et de volume par secteur sont supposés exogènes. Toutes les variables ont été centrées au préalable par année, par secteur (au niveau NES36) et par taille d'entreprise (quatre catégories).

Champ : 33 896 entreprises pérennes du secteur de l'industrie observées de 1994 à 2000.

Les deux premières colonnes représentent l'estimation de la fonction de production selon la méthode d'Arellano et Bond, après centrage par secteur et année des variables explicatives et dépendante et avec les niveaux en (t-3) et (t-2) des variables explicatives comme instruments. La troisième colonne présente les résultats de l'estimation avec tous les retards de (t-5) à (t-2) comme instruments. Dans la quatrième estimation, les variables ne sont pas centrées, mais des indicatrices pour chacune des années d'observation sont incluses parmi les régresseurs. Dans la dernière régression, enfin, on a ajouté des termes croisant capital et part des catégories de qualification dans l'emploi.

L'une des raisons invoquées par Griliches et Mairesse (1995) pour expliquer l'estimation de rendements d'échelles très bas dans la méthode d'Arellano et Bond est que, du fait des transformations faites pour corriger le problème d'endogénéité, la corrélation entre les variables

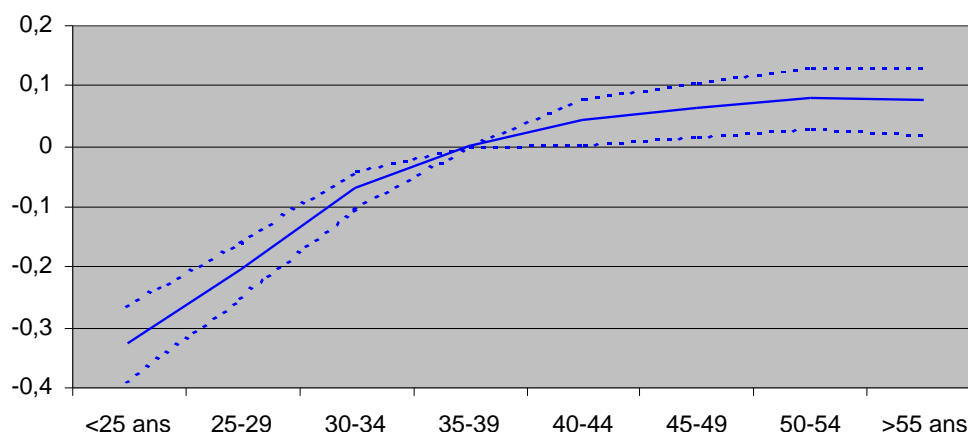
explicatives et leurs instruments est de plus en plus faible. Ceci peut exacerber d'autres problèmes, en particulier le biais en direction de zéro dû aux erreurs de mesure. Cet argument semble être justifié par nos résultats. Lorsqu'on utilise comme instruments les seuls retards en (t-2), les rendements d'échelles sont très bas, proches de 0,4. En revanche, ces rendements d'échelle estimés sont supérieures à 0,7 si l'on met deux retards dans les instruments, et même proches de 0,8 lorsqu'on utilise tous les retards de (t-5) à (t-2). Si l'on interprète l'ajout de retards parmi les instruments comme une amélioration du pouvoir prédictif de ces instruments, l'idée d'un biais d'atténuation dû à l'exacerbation des erreurs de mesure lorsque les instruments sont peu corrélés aux variables explicatives semble plausible.

Le centrage par secteur ne modifie que la productivité marginale estimée pour la part de femmes dans l'effectif. En l'absence de centrage par secteur, les femmes sont estimées significativement plus productives que les hommes, alors que leur productivité est systématiquement non-significative dans les autres cas. Par ailleurs, la décroissance apparente de la productivité avec la qualification est estimée qu'il y ait centrage ou non. Lorsqu'on ajoute des termes croisés de la part de catégorie de qualification et du capital, le coefficient de la part de très qualifiés devient positif, mais l'écart type est très élevé et ce résultat n'est pas significatif.

Les productivités marginales relatives des classes d'âge sont robustes d'une spécification à l'autre. Les coefficients estimés varient peu lorsqu'on ajoute parmi les régresseurs les parts des différentes catégories de qualification, la part de femmes ou les termes croisés, et selon que l'on centre par secteur ou non. Le profil âge-productivité induit par ces coefficients est représenté sur la figure suivante. Il dessine une relation positive et concave entre productivité et âge. La croissance de la productivité s'atténue à partir de 40 ans, pour tendre vers un niveau à peu près constant au-delà de 45 ou 50 ans.

Notons que ces résultats s'interprètent bien en ce qui concerne l'allure générale du profil âge-productivité, c'est à dire le lien positif entre l'âge et la productivité. L'interprétation des valeurs de coefficients est plus hasardeuse, d'une part du fait du biais d'atténuation discuté précédemment, et d'autre part parce qu'il faut tenir compte de la productivité marginale du travail pour calculer la productivité marginale relative d'une classe d'âge par rapport à celle des 35-39 ans (i.e. le rapport $\frac{\lambda_j}{\lambda_0}$ dans la fonction de production). Cette grandeur estimée serait donc susceptible de souffrir d'un certain nombre de biais importants.

Profil âge-productivité



Ce « profil » a été tracé par interpolation linéaire des coefficients estimés des variations des parts de chaque catégorie d'âge dans la dernière spécification du tableau précédent. La référence est la classe d'âge des 35-39 ans. Les traits en pointillés représentent les valeurs estimées des paramètres plus ou moins 1,96 écarts-types.

Par ailleurs, il nous faut rappeler qu'il ne s'agit là que du profil des productivités marginales moyennes pour chaque catégorie d'âge. Il n'est donc pas exempt d'un biais de composition, par exemple si la productivité marginale supérieure estimée pour un groupe par rapport à un autre provient du fait que ce groupe contient plus d'individus qualifiés que l'autre, et non du fait que les individus du premier groupe sont plus productifs que ceux du second, à caractéristiques plus fines égales. Par ailleurs, la productivité relative des plus de 55 ans peut traduire un effet de sélection si les salariés les moins productifs ont été exclus du marché du travail, et cela de manière plus massive dans cette tranche d'âge que dans les autres. Dans la mesure où nous estimons des fonctions de productions et ne prenons donc en compte que les individus en emploi, l'estimation d'une productivité par tranche d'âge indépendamment de l'état d'activité ou non est malheureusement impossible.

Afin de palier à cet éventuel biais de composition, nous réestimons la fonction de production en considérant des catégories plus fines de travailleurs, croisant classe d'âge et niveau de qualification. Le tableau qui suit présente les résultats de ces estimations.

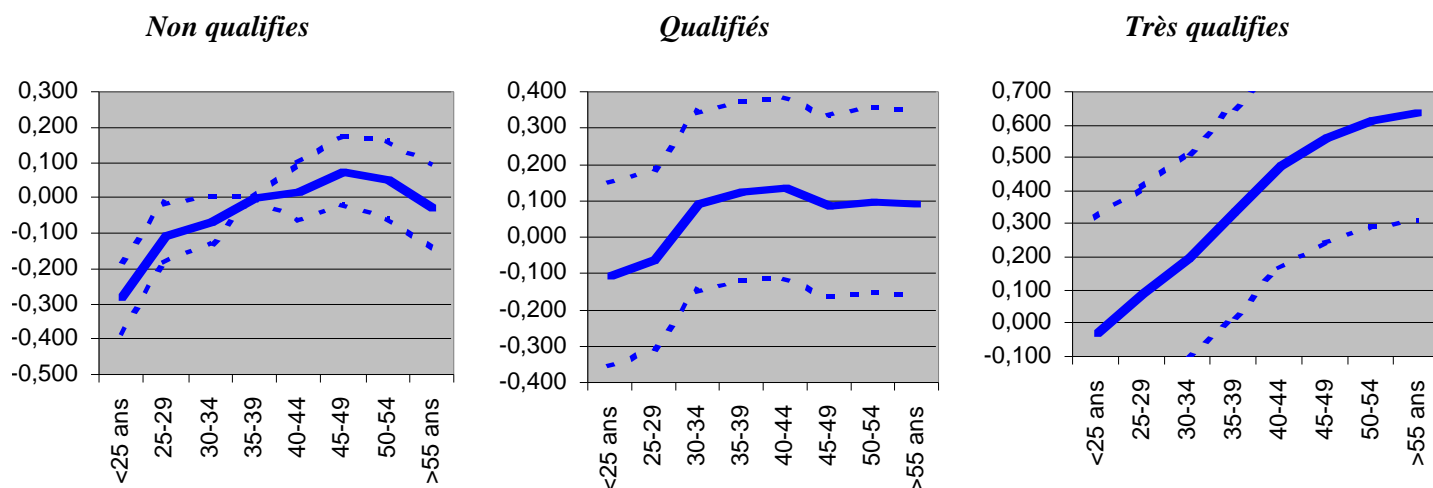
Variable	param	stde	param	stde
Capital	0,203	0,017	0,186	0,019
Nombre d'heures travaillées	0,479	0,026	0,430	0,023
D(LOGL*LOGK)			0,020	0,003
D(part(Q0)*LOGK)			0,036	0,040
D(part(Q2)*LOGK)			-0,026	0,014
D(part(Q3)*LOGK)			-0,095	0,023
Part des stagiaires et apprentis	-0,758	0,068	-0,941	0,291

Non-qualifiés				
<25 ans	-0,388	0,068	-0,280	0,052
25-29	-0,168	0,056	-0,103	0,041
30-34	-0,082	0,041	-0,065	0,033
35-39	0,000	0,000	0,000	0,000
40-44	0,028	0,068	0,016	0,041
45-49	0,097	0,069	0,075	0,049
50-54	0,085	0,070	0,050	0,055
>55 ans	0,004	0,069	-0,026	0,059
Qualifiés				
<25 ans	-0,360	0,067	-0,104	0,128
25-29	-0,293	0,062	-0,062	0,126
30-34	-0,100	0,060	0,093	0,125
35-39	-0,042	0,060	0,126	0,125
40-44	-0,012	0,058	0,135	0,126
45-49	-0,050	0,059	0,087	0,127
50-54	-0,037	0,063	0,101	0,130
>55 ans	-0,048	0,065	0,094	0,129
Très qualifiés				
<25 ans	-0,751	0,103	-0,030	0,179
25-29	-0,609	0,084	0,093	0,164
30-34	-0,438	0,078	0,201	0,160
35-39	-0,270	0,076	0,337	0,161
40-44	-0,104	0,074	0,477	0,160
45-49	0,001	0,076	0,564	0,163
50-54	0,061	0,079	0,617	0,165
>55 ans	0,074	0,080	0,637	0,167
Sexe				
Femmes	-0,052	0,068	-0,030	0,064

Statistique de Sargan	394	438
Degrés de liberté	216	248
P-valeur	0	0

Comme dans le cas précédent, la productivité relative des différentes catégories de qualification semble biaisée. A classe d'âge identique, les écarts entre niveau de qualification changent considérablement selon qu'apparaissent ou non les termes croisant part des qualifications et capital. En revanche, le profil par âge de la productivité au sein de chaque catégorie de qualification est robuste d'une spécification à l'autre. En particulier, l'ajout de termes croisés du capital et des parts de qualification « décale » les profils âge-productivité des qualifiés et très-qualifiés, avec des écarts-types maintenant trois à quatre fois plus gros, mais ne déforme pas le profil.

Profil âge-productivité par catégorie de qualification dans l'industrie



Le profil âge-productivité est le plus pentu parmi les très qualifiés. Dans cette catégorie, la productivité croît avec l'âge tout au long de la vie active, même si cette croissance est plus faible aux tranches d'âge élevées. Parmi les qualifiés, en revanche, la productivité semble croître jusque vers 35-40 ans, puis rester à peu près constante. Enfin, le profil par âge de la productivité paraît un peu plus pentu pour les non-qualifiés que pour les qualifiés. La productivité s'accroît jusque 40-45 ans. Le profil semble par ailleurs indiquer une diminution de la productivité après 55 ans, mais la différence n'est pas significative et il est difficile de conclure sur ce point.

5. Y a-t-il adéquation salaire productivité ?

Profils âge-salaire et âge-productivité

Le tableau suivant présente les résultats de l'estimation de l'équation de salaire. On a présenté en troisième colonne les productivités marginales relatives des classes d'âge estimées par la fonction de production i.e. le rapport des coefficients $\frac{\hat{\gamma}_i}{\hat{\alpha}}$. Les résultats semblent indiquer nettement que les salaires croissent moins vite que la productivité.

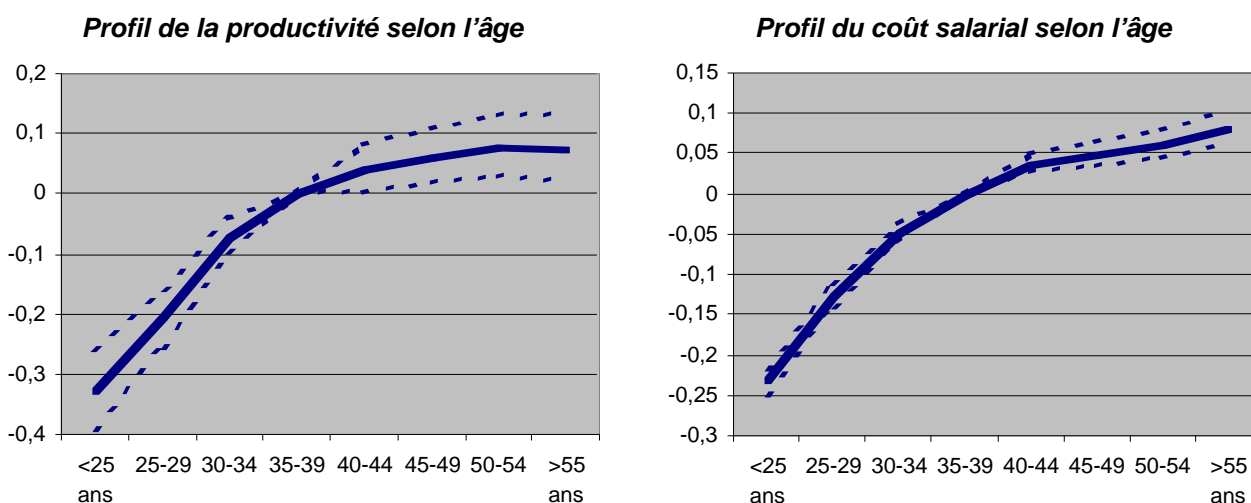
	Param	Std	Productivité estimée
	w_j		$\frac{\hat{\gamma}_i}{\hat{\alpha}}$
	w_0		$\hat{\alpha}$
<25 ans	-0,232	0,008	-0,602
25-29	-0,127	0,006	-0,379
30-34	-0,048	0,005	-0,133
35-39	0,000	0,000	0,000
40-44	0,037	0,006	0,075

45-49	0,050	0,007	0,112
50-54	0,061	0,009	0,146
>55 ans	0,082	0,011	0,138
Stagiaires, etc	-0,359	0,010	-0,861
Non qualifiés	-0,033	0,003	0,068
Très qualifiés	0,169	0,007	-0,388
Femmes	-0,095	0,012	-0,084

Champ : Estimation de l'équation de salaire en différence première linéarisée (Cf partie 2) sur 33 896 entreprises pérennes du secteur de l'industrie observées de 1996 à 2000. Les résultats sont centrés par secteur en NES36 et par année. Les productivités relatives sont estimées par la fonction de production avec tous les retards de $t-5$ à $t-2$ comme instruments.

Cependant, comme nous le soulignons dans la discussion théorique, les productivités relatives estimées sont biaisées si le coefficient du travail $\hat{\alpha}$ est biaisé. La croissance apparente nettement plus rapide pour la productivité relative que pour le coût salarial relatif pourrait donc traduire un biais négatif sur ce coefficient $\hat{\alpha}$ plutôt qu'un « rabais relatif » décroissant avec l'âge. Ce résultat semblerait donc confirmer que nous n'estimons vraiment la productivité qu'à un facteur multiplicatif près.

Si l'on compare la forme, et non plus les niveaux, des profils âge-salaire et âge-productivité ci-dessous, la croissance du salaire avec l'âge semble avoir une allure similaire à la croissance de la productivité avec l'âge. Les profils ne paraissent différents que pour le groupe des plus de 55 ans, plus payés sans être estimés plus productifs que les 50-54 ans.



Estimation du « rabais relatif »

Le tableau suivant représente les résultats de la fonction de production en remplaçant le nombre d'heures travaillées et la part des catégories dans le travail par le coût salarial et la part des catégories dans ce coût, selon la méthode développée par Crépon et al (2002).

Variable	Param	Stde	Param	Stde
Travail	0,244	0,018	0,150	0,012

Coût salarial	0,467	0,027	0,814	0,023
source :	(DADS)		(BRN)	
Age				
<25 ans	-0,225	0,040	-0,016	0,033
25-29	-0,156	0,028	0,005	0,024
30-34	-0,052	0,015	0,014	0,013
35-39	0,000	réf	0,000	réf
40-44	0,042	0,015	-0,009	0,016
45-49	0,082	0,018	-0,027	0,019
50-54	0,106	0,022	-0,037	0,021
>55 ans	0,098	0,024	-0,051	0,024
Niveau de qualification				
Stagiaires, apprentis	-0,409	0,094	-0,095	0,076
Non-qualifiés	0,061	0,028	0,088	0,025
Très qualifiés	-0,277	0,040	-0,126	0,034
Sexe				
Femme	0,051	0,076	0,096	0,063
Statistique de Sargan	403,06		246,2	
Degrés de liberté	169		169	
P-valeur	0		0	

Le résultat n'est malheureusement pas robuste au choix de la source utilisée pour le coût salarial total³. Lorsque ce coût est pris dans les DADS, on obtient un profil du « markdown », c'est à dire du rapport productivité sur salaire relatif, similaire à celui de la productivité, sans qu'il n'apparaisse d'explication très claire à cela. Au contraire, si le salaire total est mesuré dans les BRN, on estime un « markdown » légèrement décroissant avec l'âge, mais sans que les différences soient significatives. Le seul coefficient significatif est celui de la classe d'âge des plus de 55 ans, qui aurait donc bien une productivité trop faible vis-à-vis de leur salaire par rapport aux autres catégories d'âge.

6. Conclusion

A partir d'un appariement DADS-BRN couvrant plus de 33 000 entreprises pérennes des secteurs industriels entre 1994 et 2000, nous estimons des fonctions de productivité selon une méthodologie similaire à celle qui est développée dans Hellerstein, Neumark, Troske (1999) et Crépon, Deniau, Perez-Duarte (2001).

En comparant plusieurs méthodes économétriques, il semble que ces estimations sont susceptibles de souffrir d'un biais important du fait de l'endogénéité de la structure par âge des entreprises. Dans la mesure où les entreprises répondent à une demande en hausse en embauchant principalement des travailleurs jeunes et à une demande en baisse en n'embauchant peu ou plus, ce qui conduit à un vieillissement de la main d'œuvre en place, le biais d'endogénéité se traduit par une

³ Notons qu'il l'était dans le cas de la productivité. Qu'on choisisse comme mesure du travail le nombre d'heures travaillées dans les DADS ou l'effectif dans les BRN, on estime à chaque fois un profil âge-productivité croissant et concave.

productivité marginale estimée trop élevée pour les classes d'âge jeunes et trop faible pour les classes plus âgées. Cette remarque nous conduit à invalider les méthodes d'estimation qui font l'hypothèse d'une exogénéité des facteurs de production. Dans la mesure où un choc peut se traduire de manière durable sur le niveau des facteurs de production d'une entreprise, nous renonçons également à la méthode d'Arellano et Bover (1995) d'instrumentation des niveaux par les différences passées.

Nous procédons donc à des estimations de fonctions de production par la méthode d'Arellano et Bond (1991), en instrumentant les évolutions des variables par leurs niveaux passés. Cette méthode donne généralement des résultats insatisfaisants, due principalement à une faible corrélation entre les instruments et les variables explicatives. Dans le cas de la structure par âge, le vieillissement naturel de la main d'œuvre fournit néanmoins une source de variation exogène, si bien que la structure par âge d'une entreprise est un bon instrument de son évolution future.

En appliquant cette méthode et en considérant des tranches d'âge de cinq ans en cinq ans, nous estimons un profil de la productivité croissant et concave en fonction de l'âge. Cette productivité croît moins rapidement à partir de 40 ans et stagne au delà de 50 ans, mais nos estimations semblent invalider l'idée d'une productivité décroissante aux âges élevés. Ce résultat est par ailleurs robuste, dans la mesure où la forme de ce profil reste la même sous diverses spécifications. Une décomposition plus fine, croisant classe d'âge et catégorie de qualification, nous permet d'affiner ce résultat. La croissance de la productivité est la plus forte chez les très qualifiés, et reste positive tout au long de la vie active, même si la croissance est plus faible aux âges élevés. En revanche, la productivité des qualifiés croît jusque vers 40 ans, puis reste à peu près constante. Enfin, la productivité des non-qualifiés croît un peu plus rapidement que celle des qualifiés aux âges jeunes, mais elle semble décliner après 50 ans, même si cette « diminution » n'est pas significative.

En ce qui concerne l'adéquation entre salaire et productivité relative, il est difficile de conclure, du fait d'un manque de robustesse des résultats. Lorsque le coût salarial total est mesuré dans les BRN, le salaire semble croître un peu plus rapidement que la productivité, mais cela ne se traduit de manière significative que chez les plus de 55 ans.

7. Bibliographie

Arellano, Manuel, et Bond, Stephen, « Some Tests of Specification for Panel Data : Monte Carlo Evidence and an Application to Employment Equations », *Review of Economic Studies*, 58 (1991) : 277-97

Arellano, M., et Bover, O., « Another look at the Instrumental-Variable Estimation of Error-Components Models », *Journal of Econometrics* 68 (1995): 29-51

Crepon, Bruno, Deniau, Nicolas, et Perez-Duarte, Sébastien, « Productivité et Salaire des travailleurs âgés », in *Congrès des Economistes Belges de Langue Française*, Commission n°4, Chapitre 10

Crepon, Bruno, Deniau, Nicolas, et Perez-Duarte, Sébastien, « Wages, Productivity, and Worker Characteristics: A French Perspective », miméo

Griliches, Zvi, et Mairesse, Jacques. "Production Functions: The Search for Identification." Working paper 9730, CREST (1997).

Hægeland, Torbjørn, et Klette, Tor Jacob. "Do Higher Wages Reflect Higher Productivity? Education, Gender and Experience Premiums in a Matched Plant-Worker Data Set." In *The Creation and Analysis of Employer-Employee Matched Data*, edited by J. Haltiwanger, J. Lane, J. R. Spletzer, J. Theeuwes, and K. Troske, Amsterdam: North Holland, 1999.

Hellerstein, Judith K.; Neumark, David; et Troske, Kenneth R. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 17 (1999): 409–46.

Annexe 1 : Statistiques descriptives

Tableau 1.1 : Distribution des entreprises par taille

Taille	Industrie				Construction			
	Nb de firmes	Pourcent.	Nb employés	Pourcent.	Nb de firmes	Pourcent.	Nb employés	Pourcent.
Moins de 5	6 219	18,4	26 570	1,3	5 909	26,4	22 876	4,7
5 à 19 employés	13 788	40,7	159 993	7,6	11 210	50,1	124 097	25,3
20 à 99 employés	10 638	31,4	466 378	22,1	4 725	21,1	182 036	37,1
100 à 499 employés	2 704	8,0	575 094	27,2	471	2,1	85 484	17,4
Plus de 500	547	1,6	885 446	41,9	72	0,3	75 865	15,5
Total	33 896	100,0	2 113 481	100,0	22 387	100,0	490 358	100,0
Taille	Commerce				Services			
	Nb de firmes	Pourcent.	Nb employés	Pourcent.	Nb de firmes	Pourcent.	Nb employés	Pourcent.
Moins de 5	30 782	44,8	108 766	8,4	19 273	37,2	73 858	3,5
5 à 19 employés	27 550	40,1	267 340	20,7	21 263	41,0	227 105	10,8
20 à 99 employés	9 162	13,3	364 160	28,1	8 980	17,3	383 194	18,3
100 à 499 employés	1 122	1,6	209 245	16,2	2 009	3,9	400 722	19,1
Plus de 500	164	0,2	344 987	26,7	336	0,7	1 013 389	48,3
Total	68 780	100,0	1 294 498	100,0	51 861	100,0	2 098 268	100,0

Tableau 1.2 : Distribution des entreprises par année de création

Année de création	Industrie				Construction			
	Nb de firmes	Pourcent.	Nb employés	Pourcent.	Nb de firmes	Pourcent.	Nb employés	Pourcent.
Avant 1960	3 442	10,2	481 461	22,8	<i>To be completed</i>			
1960 à 1975	6 925	20,4	442 669	21,0				
1975 à 1990	19 329	57,0	1 036 177	49,0				
Après 1990	4 200	12,4	153 174	7,3				
Total	33 896	100,0	2 113 481	100,0				
Année de création	Commerce				Services			
	Nb de firmes	Pourcent.	Nb employés	Pourcent.	Nb de firmes	Pourcent.	Nb employés	Pourcent.
Avant 1960	3 558	5,2	164 831	12,7	2 101	4,1	277 313	13,2
1960 à 1975	11 549	16,8	286 505	22,1	6 176	11,9	285 905	13,6
1975 à 1990	41 664	60,6	695 479	53,7	32 584	62,8	1 061 957	50,6
Après 1990	12 009	17,5	147 683	11,4	11 000	21,2	473 093	22,6
Total	68 780	100,0	1 294 498	100,0	51 861	100,0	2 098 268	100,0

Annexe 2 : Profil âge-productivité pour les secteurs de la construction, du commerce et des services

Variable	Construction		Commerce		Services	
	param	<i>stde</i>	param	<i>stde</i>	param	<i>stde</i>
Capital	0,067	0,009	0,243	0,016	0,210	0,018
Nombre d'heures travaillées	0,598	0,032	0,539	0,025	0,512	0,029
<25 ans	-0,316	0,032	-0,326	0,032	-0,365	0,036
25-29	-0,214	0,026	-0,212	0,024	-0,236	0,028
30-34	-0,112	0,016	-0,076	0,015	-0,085	0,017
35-39	0,000	0,000	0,000	0,000	0,000	0,000
40-44	0,044	0,019	0,043	0,020	0,069	0,021
45-49	0,127	0,024	0,068	0,023	0,105	0,025
50-54	0,194	0,030	0,088	0,026	0,131	0,029
>55 ans	0,204	0,033	0,083	0,029	0,127	0,032
Stagiaires, etc	-0,400	0,055	-0,474	0,050	-0,424	0,053
Non qualifiés	0,002	0,032	0,048	0,026	0,056	0,026
Très qualifiés	-0,217	0,053	-0,210	0,047	-0,253	0,049
Femmes	-0,373	0,102	0,011	0,066	-0,045	0,071
Observations	22 837		68 780		51 861	
Sargan	216,6		380,9		257,3	
Degrés de liberté	169		169		169	
P-valeur	<0,0001		<0,0001		<0,0001	