UNIVERSITÉ D'ÉVRY-VAL-D'ESSONNE

DÉPARTEMENT DE SCIENCES ECONOMIQUES

THÈSE

pour le Doctorat en Sciences Économiques
(arrêté du 30 mars 1992, arrêté du 18 janvier 1994)
présentée et soutenue publiquement par

Stéphane ADJEMIAN

le 11 décembre 2002

DIVERGENCE DES NATIONS ET PROGRÈS TECHNIQUE

DIRECTEUR

Monsieur Jérôme Glachant, Professeur à l'Université d'Évry

JURY

Monsieur Alain Desdoigts, Professeur à l'Université de Dijon
Monsieur Patrick Fève, Professeur à l'Université de Toulouse I
Monsieur Stéphane Gregoir, Administrateur INSEE
Monsieur Jérôme Glachant, Professeur à l'Université d'Évry
Monsieur Kristian Kerstens, Chargé de recherche CNRS
Monsieur Ferhat Mihoubi, Professeur à l'université d'Évry

Remerciements

Même si cette thèse est le fruit d'un travail solitaire, beaucoup de personnes ont directement ou indirectement contribué à sa rédaction. J'écris ces quelques lignes au dernier moment, il se peut que j'oublie de remercier quelques personnes. J'espère qu'elles n'y verront pas une offense.

Jérôme Glachant est la première personne que je souhaite remercier. Avant d'être mon directeur de thèse, il fût mon professeur d'économie préféré à l'université d'Évry. Je garde un souvenir, parmi d'autres, impérissable du TD de macro-économie qu'il dispensait lorsque j'étais petit étudiant en licence d'économétrie. Ce premier contact fût déterminant dans le choix de mon mémoire de maîtrise (déjà sur les tests de convergence, sujet pour lequel je n'avais alors a priori aucun intérêt). Depuis cette époque il a su me laisser totalement libre, tout en étant toujours présent pour m'aider et me guider. Je ne sais pas comment lui témoigner mon admiration sans tomber dans un pathos ridicule...

Stéphane Gregoir est la deuxième personne que je souhaite remercier. Sans son merveilleux cours de Séries Temporelles, j'aurais pu terminer mon second cycle d'économétrie sans témoigner le moindre intérêt pour l'économétrie ! Même si je n'ai pas compris grand chose à l'époque, je trouvais son cours mystérieux et fascinant, cela suffit parfois à motiver les étudiants. Plus tard il fût toujours présent pour répondre à mes nombreuses questions ou discuter (lorsque j'ai osé le déranger).

J'ai bénéficié du soutien constant de Ferhat Mihoubi depuis mon second cycle. Je tiens vivement à l'en remercier car rien ne l'obligeait. Ferhat a eu la gentillesse de m'accorder sa confiance en me donnant la responsabilité du cours de séries temporelles dans son IUP d'Éco–Stat. Je dois pour cela aussi le remercier, car cette charge m'a permis de faire de grands progrès (même si elle a monopolisé une partie de mes ressources).
Les beaux graphiques d'Alain Desdoigts m'ont motivé pour me mettre au non paramétrique. Alain a su me guider dans cette découverte.
Michel Guillard a bien voulu m'accepter au sein de l'EPEE, où j'ai bénéficié de très bonnes conditions matérielles pour travailler (j'écris ces quelques lignes sur mon tout nouveau super–PC avec écran plat...).
J'ai bénéficié des remarques de quelques personnes : Alain, Ferhat, Fabrice Gilles, Kristiaan Kerstens et Christophe Tavéra. Je regrette de n'avoir pu incorporer toutes leurs remarques dans mon travail.

Heureusement ces quatres années ne furent pas seulement consacrées à ma thèse. Je tiens vivement à remercier Sylvie Lefranc et surtout Ludovic Aubert de m'avoir offert des escapades intellectuelles, en travaillant sur la politique monétaire optimale. Je dois aussi remercier de nombreux étudiants rencontrés durant ces quatres années, l'enseignement est une véritable récréation... Même si je ne parviens toujours pas à comprendre pourquoi ils ne tombent pas sous le charme de l'asymptotique des processus non stationnaires.

Finalement, je dois dire un grand merci à mes parents qui ont eu beaucoup de patience.

Évry, le 17 novembre 2002

J'ai utilisé les logiciels XploRe 4.2, Ox, MatLaB, Gauss et les langages Fortran (le compilateur GNU g77) et C++ (le compilateur GNU gcc) pour les estimations, simulations et graphiques. La thèse est composée avec la distribution MiKTeX de LaTeX.

Étant donné le grand nombre de graphiques et tableaux, j'ai préféré disposer ceux-ci à la fin des chapitres. Lorsque j'ai jugé que cela pouvait grandement faciliter la lecture, j'ai ramené les graphiques et tableaux dans le texte. La version *pdf* de la thèse est munie de liens hyper texte, afin de faciliter la lecture. Les annexes techniques sont spécifiques à chaque chapitre.

Évry, le 15 novembre 2002
stephane.adjemian@eco.univ-evry.fr

Table des matières

“*There is scarcely any inquiry more curious, or, from its importance, more worthy of attention, than that which traces the causes which practically check the progress of wealth in different countries, and stop it, or make it proceed very slowly while the power of production remains comparatively undiminished, or at least would furnish the means of a great and abundant increase of produce and population.*”

THOMAS R. MALTHUS, in Principles of Political Economy (1820), Chapitre VII.

Introduction générale

> "*The newspaper is filled with accounts of monthly changes in industrial production and retail sales. But these short run changes have a relatively minor impact on economic well being . Why GDP rose or fell a few percent over the last three months can be an intriguing question. Even more significant, however, is why United States is so much richer than Nigeria or why growth in U.S. incomes over past quarter century has been slower than growth over the previous quarter century*". Gregory Mankiw, préface de Barro et Sala-i-Martin [16].

La citation de Malthus placée en exergue de cette thèse montre que la diversité des performances des nations en termes de croissance est un problème qui préoccupe les économistes depuis quelques siècles. Même si l'intérêt pour cette question a connu des hauts et des bas, il semble incontestable que ces performances soient cruciales en termes de bien être. Lucas [93], à l'aide d'exercices numériques, conclut de la façon suivante,

> "(...) *economic instability at the level we have experienced since the Second World War is a minor problem even relative to historically experienced inflation and certainly relative to the costs of modestly reduced rates of economic growth.*" Lucas [93, pages 30-31].

La valeur ajoutée, en termes de bien être social, d'une recherche sur les sources l'hétérogénéité des dynamiques de croissance est potentiellement très substantielle. Malgré cet intérêt indéniable, ce thème ne prit l'avant scène qu'au milieu des années 80. Il est presque d'usage d'entamer tout livre sur la croissance en notant que la discipline de la croissance économique s'est récemment éveillée d'un sommeil de 20 ans. Barro et Sala-i-Martin n'échappent pas à la règle,

> *"Economists have, in some sense, always known that growth is important. Yet at the core of the discipline, the study of economic growth languished after the late 1960s. Then, after a lapse of nearly two decades, this research became vigorous again in the mid-1980s."* Barro-Sala-i-Martin [16, page xvii].

Cette rhétorique cache un changement profond. Le regain d'intérêt pour l'économie de la croissance se manifeste en une nouvelle perspective. Les modèles de croissance des années 60 ne cherchent pas à expliquer la diversité des dynamiques de croissance. Le modèle de Solow [140] s'attache à expliquer la croissance à moyen terme d'une économie, mais ne propose aucune explication des disparités de croissance. Plus tard, en situant son travail par rapport aux faits stylisés de Kaldor, Solow [142] note que,

> "*The remaining stylized facts are of a different kind, and will concern me less, because they relate more to comparisons between different economies than to the course of events within any one economy.*" Solow [142, page 3].

On peut ainsi dire que Barro et Sala-i-Martin [15] ou Mankiw, Romer et Weil [95] détournent le modèle de Solow, en essayant de montrer qu'il peut tout aussi bien expliquer la diversité des performances en termes de croissance.

L'origine de ce changement de perspective est double. Premièrement, il n'aurait pu s'exercer sans la constitution d'une base de données internationales : les *Penn World Tables* de Summers et Heston [146]. Cette base de données fournit des informations sur le produit par tête, et ses décompositions. La correction pour la parité des pouvoirs d'achat et la considération d'indices multilatéraux autorisent la comparaison des performances dans le temps et dans la coupe des nations[1]. La méthodologie consiste essentiellement en la construction d'indices des prix corrigés de la parité des pouvoirs d'achat. La construction de tels indices réclame la collecte d'informations,

[1] Pour une présentation de la méthode, on peut se reporter à Kravis, Heston et Summers [83]. Pour une présentation critique, on peut lire Dowrick et Quiggin [41].

à une échelle relativement fine, sur les prix de centaines de biens[2] pour un maximum de nations en des années régulièrement espacées (cinq ans). Ces exercices de collecte, à la base des *Penn World Tables*, sont les *ICP benchmark studies* (*ICP, pour International Comparison Project*). Entre 1970 et 1990, 90 nations ont participé à au moins un de ces projets. Une agrégation des prix collectés permet d'obtenir des indices des prix puis de quantités comparables dans le temps et l'espace. Les *Penn World Tables* souffrent de plusieurs défauts. (*i*) La collecte des prix sur une large coupe de nations n'est pas sans poser problème. Les comptabilités nationales ne sont pas toutes de qualités égales. La non participation régulière de nombreuses nations aux *ICP benchmark studies* oblige la mise en œuvre d'interpolations pour obtenir les indices de la PPA sur une large coupe de nations. Les données produites sont donc de qualités inégales. (*ii*) Il n'est pas évident que les paniers de biens, censés être identiques dans la coupe, soient réellement comparables. Summers et Heston reconnaissent que,

> "*A physical potato may be virtually the same everywhere, but can adequate allowance be made for the difference between potatoes purchased in air conditioned supermarkets in rich countries and those purchased at roadside stands in poor countries ?*" Summers et Heston [147, page 22]

(*iii*) L'agrégation des prix et quantités n'est pas un problème trivial, *ie* le choix d'un indice n'est pas neutre et repose sur une axiomatique des préférences. Il n'est pas sûr que l'existence de préférences communes au niveau mondial, implicite dans les comparaisons de produit fournies par les *Penn World Tables*, soit une hypothèse légitime[3]. (*iv*) Les évaluations du produit sont obtenues du côté des dépenses et non pas en sommant les valeurs ajoutées. On peut se demander dans quelle mesure il est légitime d'utiliser ces données pour comparer des productivités. (*v*) Les données relatives à l'emploi sont, de l'aveu même des auteurs, peu fiables (surtout pour les nations en dehors de l'OCDE). Malgré ces limites, et la liste n'est pas exhaustive, les *Penn World Tables* constituent aujourd'hui la base de données interna-

[2]Ces biens sont censés être homogènes dans la coupe des nations.

[3]Marris [96] a mis en avant ce point. Toutefois on ne dispose pas aujourd'hui de résultats empiriques pour rejeter l'hypothèse de préférences communes dans la coupe. Voir, par exemple, Dowrick et Quiggin [40] qui concluent "*The data, however, are quite consistent with the hypothesis that the choice between sushi and schnitzel is driven by relative prices*".

tionales la plus intéressante.

Deuxièmement, ce changement de perspective a participé au débat entre plusieurs théories "*alternatives*" de la croissance. Ainsi Romer déclare,

> "*Both Robert Lucas (1988) and I (Romer, 1986) cited the failure of cross-country convergence to motivate models of growth that drop the two central assumptions of the neoclassical model : that technological change is exogenous and that the same technological opportunities are available in all countries of the world.*" Romer [126, page 4]

La question de la convergence des économies — une réduction du problème de la comparaison des performances nationales — devient alors l'instrument discriminant entre plusieurs théories de la croissance. De nombreux articles empiriques sur la convergence motivent le sujet comme un test de la croissance exogène contre la croissance endogène[4]. Le problème de cette présentation est que ces "*alternatives*" ne sont pas nécessairement exclusives du point de vue de la convergence des nations. Le modèle néo-classique de croissance exogène ne prédit pas que les économies convergent absolument et des modèles de croissance endogène peuvent prédire l'existence d'une dynamique de transition où une économie croît d'autant plus vite qu'elle est éloignée de sa cible de long terme. Ainsi Pack souligne que,

> "*Tests of the neoclassical growth framework will not, in and of themselves, prove very much about the value of endogenous growth theories.*" Pack [105, page 70]

et finalement Romer, qui instrumentalise la question de la convergence dès 1986, reconnaît que,

> "*(...) the convergence controversy captures only part of what endogenous growth has been all about. It may encompass a large fraction of the recently published papers, but it nevertheless represents a digression from the main story behind endogenous growth theory.*" Romer [126, page 11]

Dans cette thèse on ne cherche pas à discriminer entre ces théories de la croissance. La question principale qui nous intéresse ici est de savoir dans

[4]Voir par exemple les articles d'Evans [49], [51].

quelles mesures la diversité des performances observées s'explique par l'accumulation de facteurs ou par la technologie.

Lorsque l'on considère le modèle de Solow [140] pour rendre compte de la diversité des dynamiques de croissance (comme Mankiw, Romer et Weil [95]), on suppose que la technologie est commune et qu'elle se diffuse instantanément. Dans ce cadre, la diversité des performances des nations ne peut s'expliquer que par l'accumulation du capital physique (et éventuellement du capital humain). L'hypothèse de rendements décroissants du capital physique induit un ralentissement de la croissance lorsque l'économie se rapproche de sa cible de long terme. Ainsi, le taux de croissance d'une économie est d'autant plus important qu'elle est éloignée de son état stationnaire. Toutes choses égales par ailleurs, le modèle prédit alors une convergence des nations. On dira que la convergence est néo-classique, si elle s'explique par ce mécanisme.

D'autres économistes (comme Baumol [17]), sans remettre en cause l'hypothèse d'une convergence des nations, expliquent la diversité des performances en termes de croissance en appelant un autre mécanisme. Ils ne supposent pas, comme Mankiw, Romer et Weil, la diffusion instantanée de la technologie. Ainsi, certaines nations souffrent d'un retard technologique, dans le sens où elles ne peuvent pas exploiter efficacement la meilleure pratique de production. C'est alors une plus ou moins lente diffusion de la meilleure pratique qui induit une force de rappel résultant en une convergence des nations. La vitesse de cet ajustement dépend de la capacité des nations à absorber la meilleure pratique (environnements institutionnel et social, éducation de la main d'œuvre,...). On dira que la convergence (ou le rattrapage) est technologique, si elle s'explique par ce mécanisme.

Un des problèmes de la littérature empirique de la croissance est qu'elle emploie souvent les mêmes approches empiriques, en considérant des théories différentes pour les justifier. Par exemple, les régressions en coupe [5] sont utilisées pour rendre compte des convergences néo-classique et technologique. Une corrélation négative entre le taux de croissance moyen et la condition initiale du produit s'interprète pour les uns comme la manifestation du *capital*

[5]Voir plus loin pour une présentation.

deepening effect et pour les autres comme la manifestation d'un *technological catch-up*. Ces interprétations peuvent entraîner des recommendations de politiques économiques différentes (soutenir l'épargne d'un côté, réduire les freins à la diffusion des technologies de l'autre). Dès lors, il peut être intéressant de voir quel mécanisme est le plus important pour expliquer la diversité des dynamiques de croissance.

Le problème de la littérature empirique ne réside pas seulement dans la double interprétation offerte par les instruments employées. Elle tient aussi aux instruments eux-mêmes. Dans cette thèse on s'intéressera donc aussi aux méthodes utilisées pour tester la convergence. L'idée d'une convergence, même conditionnelle, des nations au niveau mondial ne va pas de soi. Les rôles respectifs des mécanismes néo-classique et technologique doivent aussi être évalués dans cette perspective. Si les nations divergent (c'est la position défendue ici) est-ce la conséquence d'un échec du mécanisme de convergence néo-classique ou d'un échec de la convergence technologique ?

Avant de définir de façon plus précise le contenu de cette thèse, il convient (*i*) de revenir sur les évidences empiriques qui motivent notre réflexion (*ii*) de rappeler les méthodes généralement utilisées pour tester la convergence des économies et (*iii*) de rappeler les méthodes traditionnellement utilisées pour évaluer le poids de la technologie dans la croissance.

Dynamique Mondiale

Avant d'entamer toute discussion sur les méthodes, il convient de jeter un oeil sur les données. Cette observation suffit pour voir que l'hypothèse de convergence des nations, au sens commun où elles devraient se rapprocher les unes des autres, est intenable. Une façon intuitive d'aborder cette question est de considérer l'évolution de la distribution du produit par tête. Nous suivons la démarche de Desdoigts [35].

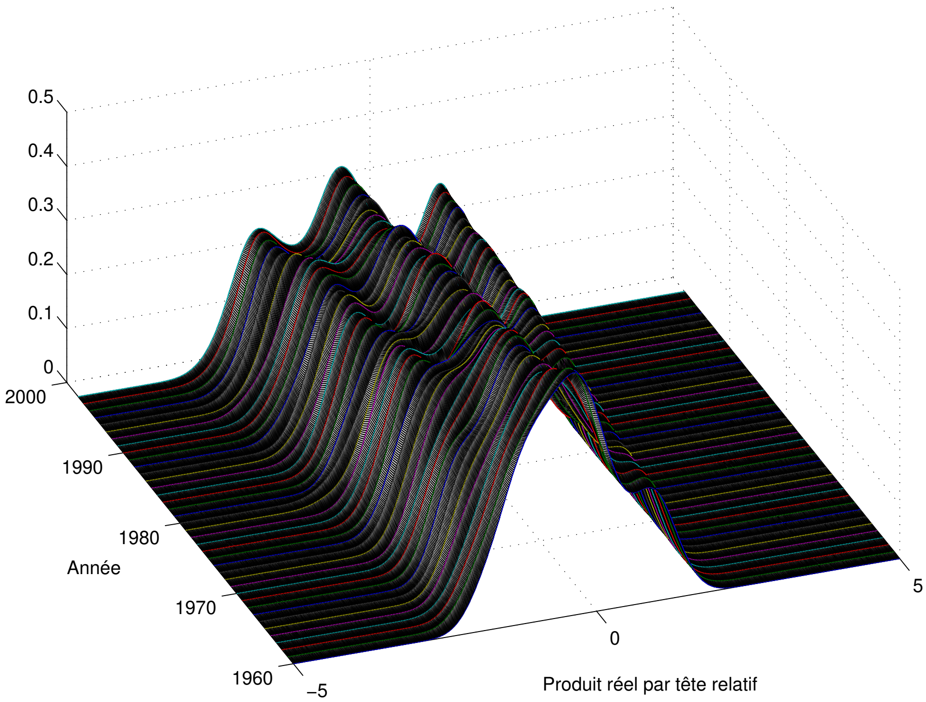

FIG. 1 – **Évolution du produit réel par tête relatif**.

La figure [1] représente l'évolution de la distribution du produit réel par tête relatif entre 1960 et 1998 pour 102 nations. Les données sont extraites des *Penn World Tables* 6[6]. Pour estimer ces distributions on a utilisé un estimateur à noyau avec un paramètre de lissage constant dans la chronologie égal à 0.2845[7]. Clairement, si la distribution du produit par tête semble unimodale en 1960, on note que la distribution se déforme pour devenir multimodale en 1998. Une façon alternative de représenter cette tendance de la distribution du produit par tête est de regarder l'évolution, de date en date, du paramètre de lissage optimal estimé.

[6]On a utilisé la variable intitulée RGDPL, voir Summers et Heston [146].

[7]Il s'agit de la moyenne arithmétique des paramètres de lissage optimaux obtenus en minimisant l'intégrale de l'erreur quadratique moyenne par une procédure de *cross-validation*.

FIG. 2 – **Évolution du paramètre de lissage optimal estimé**.

En effet, Silverman [132, page 139] montre que le nombre de modes d'une distribution estimée est une fonction décroissante du paramètre de lissage[8]. La figure [2] rend compte de cette évolution. On note que le paramètre de lissage optimal estimé diminue quasi régulièrement entre 1960 et 1980. En faisant abstraction du fait que ce graphique ne reporte que des estimations, ceci est cohérent avec l'observation précédente d'une distribution du produit par tête tendant vers une distribution multi-modale. Cette observation contredit l'idée même de convergence. Ces deux graphiques nous disent que les nations ont globalement tendance à s'éloigner les unes des autres, car la distribution du produit par tête devient multi-modale. Sous l'hypothèse de convergence, la distribution devrait au moins être stable dans la chronologie. Ici on observe que cette distribution a tendance à s'étaler et que les nations se regroupent ; on discerne trois modes en 1998. Mais la distribution du produit par tête en

[8]Il propose ainsi un test de l'uni-modalité contre la multi-modalité en utilisant le paramètre de lissage. Ce test est mis en œuvre par Bianchi [24] pour tester l'uni-modalité de la distribution du produit par tête en 1970, 1980 et 1989. Il rejette l'hypothèse d'une distribution uni-modale.

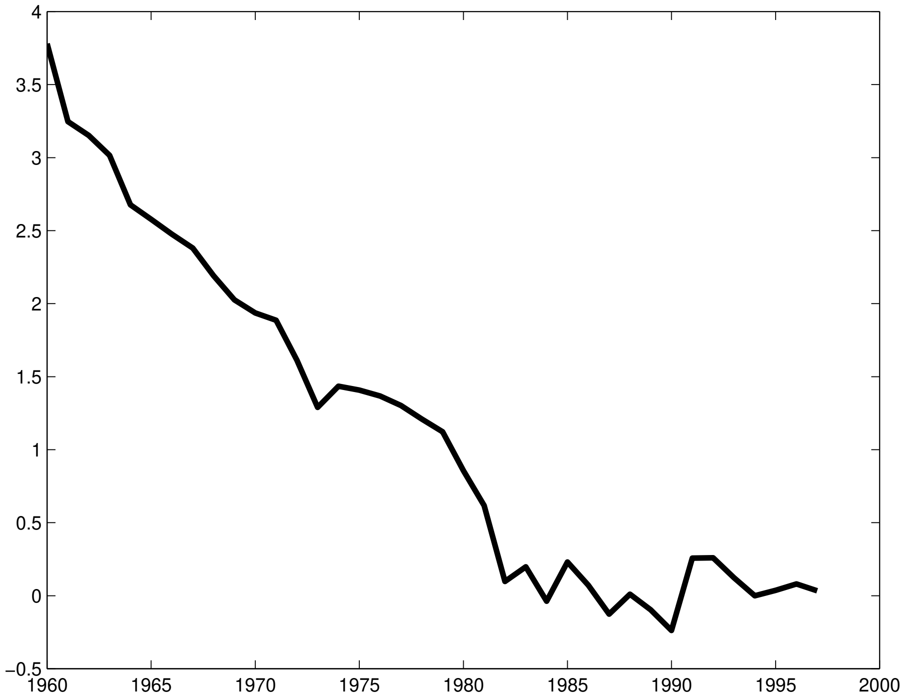

FIG. 3 – **Évolution de la statistique estimée de Li.**

1998 est-elle significativement différente de celle de 1960 ? Pour répondre à cette question on utilise la statistique de Li [91], qui permet de tester l'identité de deux distributions inconnues. La figure [3] reporte les estimations de cette statistique pour tester l'identité de la distribution du produit par tête en 1998 avec la distribution du produit par tête en $t = 1960, \ldots, 1997$. Encore une fois on observe un profil globalement décroissant. Ce test unilatéral est asymptotiquement normalement distribué[9]. Ainsi, la distribution du produit par tête en $t = 1960, \ldots, 1971$ est significativement différente de la distribution du produit par tête en 1998 au seuil de 5%.

Toutes ces évidences empiriques vont dans le même sens : *les économies divergent*. L'hétérogénéité observée des taux de croissance moyens sur la période 1960–1998, ne contribue pas à un rapprochement des nations mais à une divergence des nations.

[9]Il faut donc comparer les statistiques estimées à 1,64.

Une observation plus originale est que ce processus de divergence n'est pas invariant. En effet, un second regard sur la figure [3] montre que la statistique de Li devient relativement stable autour de zéro à partir des années 80. Durant la même période, la chute du paramètre de lissage semble bien moins affirmée (voir la figure [2]). Ces observations suggèrent que l'évolution de la distribution du produit par tête est bloquée à partir des années 80. Si on considère seulement le test d'identité des distributions de produit par tête, cet arrêt de la dynamique de la distribution du produit par tête se manifeste dès le début des années 70. Cela ne nous permet pas de conclure que le produit par tête atteint sa distribution de long terme entre 1970 et 1980, mais que, pour l'essentiel, la divergence des nations se manifeste avant 1980. On retrouvera cette instabilité de la dynamique de croissance dans les chapitres [2] et [3].

On associe régulièrement[10] l'observation de distributions multi-modales du produit par tête avec l'idée de *club* de convergence. Cette notion est clairement définie par Galor,

> "*The* club convergence hypothesis *(polarization, persistent poverty, and clustering) – per capita incomes of countries that are identical in their structural characteristics converge to one another in the long run provided that their initial conditions are similar as well.*" Galor [60, page 1056]

puis il précise dans une note de bas de page,

> "*That is, countries converge to one another if their initial conditions are in the basin of attraction of the same steady-state equilibrium.*" Galor [60, page 1056]

L'observation d'une distribution multi-modale pour le produit par tête ne prouve pas que l'on soit en présence de *clubs* de convergence. En effet la tendance vers une distribution multi-modale de la distribution du produit par tête peut disparaître lorsque l'on conditionne le produit par tête par des variables comme la proximité spatiale, l'ouverture ou encore le niveau d'éducation. Quah [123] donne de tels exemples sur les régions européennes.

[10] Voir, par exemple, la note de bas de page 3 dans Galor [60].

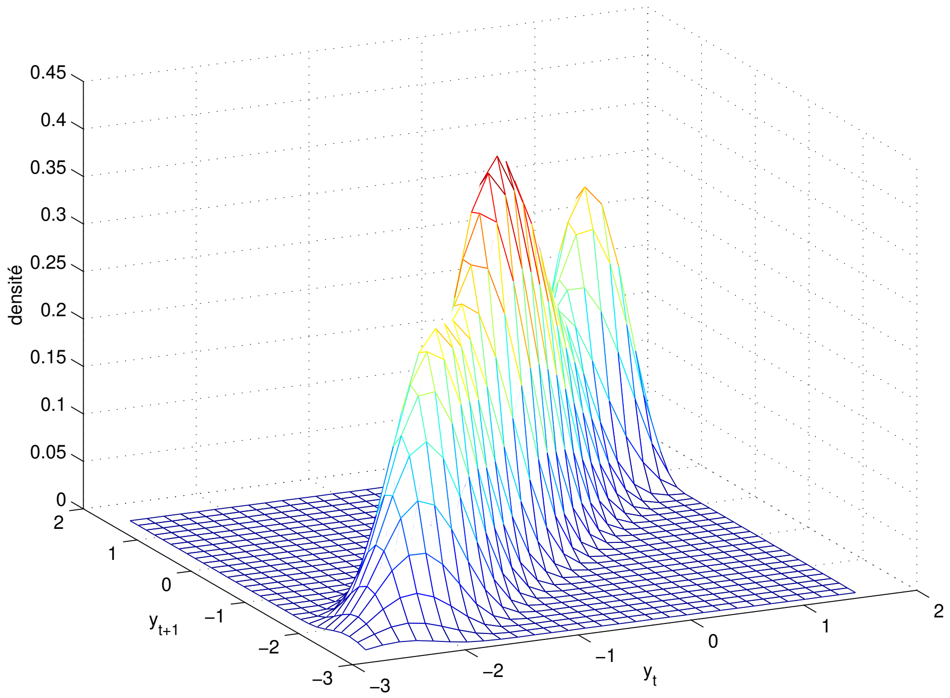

FIG. 4 – **Stochastic kernel**.

Cette démarche empirique fournit peu d'information sur les dynamiques à l'intérieur de la distribution du produit par tête. Il ne s'agit que d'estimations de la densité du produit par tête à une date donnée. Cela ne peut donner d'idée claire sur la mobilité dans la coupe des nations. Pour ce faire une idée plus précise des propriétés de la dynamique à l'intérieur de la distribution du produit par tête, on peut, en suivant Quah [120], construire des matrices de transition pour la distribution discrétisée du produit par tête, ou mieux, adopter une version continue de ces matrices de transition comme dans Desdoigts [35, page 80] ou Quah [122] : les *stochastic kernel*. Une *stochastic kernel* est une densité bivariée indiquant la probabilité pour qu'une nation, dont le produit par tête est dans un voisinage de y à la date t, ait un produit par tête dans un voisinage de y' à la date $t+T$. La figure [4] représente un estimateur de cette densité bivariée pour notre échantillon. Cette *stochastic kernel* est obtenue en considérant le produit par tête relatif à la date t contre le produit par tête relatif à la date $t+1$. On note que la masse de probabilité de cette distribution s'étend essentiellement le long de la première bissectrice. Ceci

suggère que la dynamique du produit par tête est très persistante. Une économie initialement riche (respectivement pauvre) a une grande probabilité de rester riche (respectivement pauvre). De plus, on note que la dynamique du produit par tête relatif semble non linéaire, dans le sens où cette distribution bivariée est multi-modale, et suggère à nouveau une convergence locale (c'est à dire l'existence de *clubs* de convergence). Notons toutefois que cette observation n'est pas réellement concluante, car en conditionnant le produit par tête par une variable (comme l'investissement ou l'ouverture) le caractère multi-modal de cette distribution peut disparaître — voir par exemple Proudman, Redding et Bianchi [118]. Dans ce cas, le niveau de long terme ne paraît plus conditionné par le produit par tête initial et on ne peut véritablement identifier des *clubs* de convergence.
Enfin, notons que cette approche, si elle donne plus d'information sur la dynamique du produit par tête, ne permet pas d'estimer des vitesses de convergence locales (on reviendra plus loin sur ce point en comparant cette approche avec les régressions en coupe).

Définir et tester la convergence

Les définitions de la convergence sont nombreuses. Nous présentons ici les définitions les plus répandues, associées à ce que nous appellerons l'*approche classique*[11] et la définition basée sur les propriétés des séries temporelles[12].

L'*approche classique* dispose de deux instruments pour résumer les données et conclure sur la convergence. La première consiste en l'observation de l'évolution de la variance en coupe du produit par tête. Si on observe une diminution de la variance en coupe alors les économies convergent au sens de la σ–convergence. Cette définition, si elle est intuitive, a deux limites. Premièrement, pourquoi ne s'intéresser qu'au moment d'ordre deux plutôt qu'à

[11] Sala-i-Martin [127] est l'auteur de cette appellation qui sera adoptée ici. Elle se justifie simplement par l'abondance de papiers et articles abordant la question de la convergence dans cette perspective, mais surtout parce que les démarches alternatives sont toujours introduites en opposition à cette approche.

[12] Le chapitre [2] utilise cette définition alors que le chapitre [1] se concentre essentiellement sur l'*approche classique*.

tous les moments en étudiant directement l'évolution de la distribution du produit par tête. Deuxièmement, cette approche ne donne que peu d'information sur la dynamique à l'intérieur de la distribution. Une réduction de la variance implique l'existence d'une force de rappel qui tend à rapprocher les économies. Mais une inflation de la variance en coupe est aussi compatible avec l'existence de cette force de rappel (on reviendra plus loin sur ce point).

La seconde approche repose sur une idée simple. Si les économies convergent, alors le taux de croissance d'une économie initialement plus riche doit être plus faible que le taux de croissance d'une économie initialement pauvre. Si on observe une relation décroissante entre le taux de croissance et la richesse initiale, alors on devrait conclure à la convergence des économies. L'extrême simplicité de cette approche doit, pour une bonne part, expliquer son succès.
Notons que la corrélation négative entre le taux de croissance et le niveau n'est qu'une condition nécessaire du rattrapage des nations pauvres vers les riches. Malgré cela, cette condition est aujourd'hui assimilée à une définition alternative de la convergence.

Dans sa version la plus frustre, on peut citer Baumol [17] comme un article fondateur. Il ne fournit pas de référence formelle à un modèle de croissance pour justifier son approche mais donne un cadre d'analyse empirique de la convergence sous la forme (i) d'un graphique représentant le taux de croissance contre le niveau initial[13] et (ii) d'une régression du taux de croissance sur le niveau initial et une constante. En l'absence de référence formelle à un modèle, ce travail demeure purement descriptif, mais il a le mérite d'introduire un cadre d'analyse intuitif qui deviendra le prisme commun utilisé par la profession. Si le modèle n'est pas explicitement exposé, on voit que Baumol pense le phénomène de convergence des nations comme le résultat d'un rattrapage technologique[14]. Son apport ne se limite pas à l'introduction de la β–convergence, il aborde aussi le problème de l'hétérogénéité. Il note

[13] Il considère essentiellement deux échantillons et deux variables : le produit annuel par heures travaillées de 1870 à 1979 pour un échantillon de 16 économies et le produit par tête de 1950 à 1980 pour un échantillon plus large de 72 économies.

[14] Il cite l'idée de Gerschenkron [62] sur les avantages, en termes de croissance, de la *relative backwardness*.

que si pour une large coupe d'économies on n'observe pas le phénomène de convergence, on peut toutefois extraire des sous-ensembles d'économies qui convergent (Baumol parle de *clubs* mais cela ne correspond pas à la définition donnée par Galor [60]). Dans son étude, il distingue les riches économies libérales et les économies de l'Est. Il jette ainsi les bases de la β–convergence conditionnelle, même s'il ne met pas explicitement en œuvre une régression, on peut révéler une relation décroissante entre croissance et niveau initial une fois que l'on contrôle de l'hétérogénéité caractérisant éventuellement un échantillon (par exemple lié aux institutions).

On peut directement affilier toute une part de la littérature à Baumol [17]. Une importante proportion de la littérature empirique associée à l'*approche classique* tente ainsi d'expliquer la diversité des performances nationales en termes de croissance en régressant un taux de croissance sur le niveau initial et des variables susceptibles de contrôler d'une éventuelle hétérogénéité, sans formellement rattacher ces régressions à un modèle[15]. Un article représentatif de cette version descriptive de l'*approche classique* est donné par Barro [12]. Face à une telle diversité de conditionnements possibles se pose alors la question de la robustesse des résultats obtenus. Levine et Renelt notent que

> "*There does not exist a consensus theoretical framework to guide empirical work on growth, and existing models do not completely specify the variables that should be held constant while conducting statistical inference on the relationship between growth and the variables of primary interest.*" Levine et Renelt [90, page 943]

Ils travaillent sur la base d'une cinquantaine de variables susceptibles d'expliquer l'hétérogénéité des dynamiques de croissance (par exemple des mesures d'ouverture ou des indices liés aux régimes politiques) et proposent d'évaluer la robustesse des estimations des coefficients d'intérêts (*ie* les coefficients associés au produit initial, à la part de l'investissement, au niveau du capital humain) à des changements sur l'espace de conditionnement. Il ressort de cet exercice que la convergence conditionnelle sur la période 1960-1989 est un résultat robuste (comme le lien entre croissance et investissement). Néanmoins, il convient de noter que le bien-fondé statistique de ce genre d'exercice n'est

[15]Durlauf et Quah [47] proposent un résumé des différentes variables considérées dans cette littérature pour augmenter les régressions en coupe.

pas des plus clairs. Par exemple, Sala-i-Martin [128] qualifie une variable de robuste si elle apparaît significativement dans plus de 95% des régressions[16]. L'idée d'utiliser la significativité d'un coefficient pour sélectionner un modèle est en soit discutable. Comme le note Durlauf [44], cela pourrait conduire à rejeter un ensemble de variables à cause d'un problème de colinéarité alors que cette exclusion pourrait fortement dégrader le pouvoir explicatif du modèle.

Il faut attendre Barro et Sala-i-Martin [14], [15] et Mankiw, Romer et Weil [95] pour voir se développer un usage moins descriptif des régressions en coupe. Barro et Sala-i-Martin (BSM, par la suite) proposent une interprétation structurelle des paramètres de la régression en coupe, basée sur un modèle de croissance optimale néo-classique. Mankiw Romer et Weil (MRW, par la suite) proposent un modèle pour justifier l'introduction du capital humain dans une régression en coupe. Dans BSM, le paramètre associé à la condition initiale s'interprète alors comme une *proxy* du niveau initial de la variable d'état du modèle, le stock de capital physique.

Les observations empiriques sont généralement défavorables à la β–convergence absolue, sauf sur des données régionales, mais un grand nombre d'articles concluent à la β–convergence conditionnelle[17]. L'*approche classique* est la démarche empirique la plus employée, sa simplicité explique sûrement pour une bonne part son succès en termes de publications. Certains voient un signe de la robustesse de cette démarche dans sa capacité à produire, pour divers échantillons, quelques régularités empiriques. La plus célébrée est la mesure d'une vitesse de convergence se situant généralement autour de 2%. La simplicité et le succès indéniables de cette démarche empirique ne prouvent pas qu'elle soit la plus fiable. Une façon de voir dans quelle mesure cette démarche peut nous apprendre quelque chose sur la convergence est de définir le processus générateur des données cohérent avec cette démarche. Si le processus générateur des données pose des hypothèses trop restrictives, les régressions en coupe ne peuvent constituer une méthodologie intéressante.

[16]Pour chaque régression les autres variables explicatives sont changées.

[17]Voir Sala-i-Martin [127] pour une revue des résultats sur données internationales et régionales.

Ce cadre idéal (*ie* le processus générateur des données) est défini par De la Fuente [33] et Evans [49]. On suppose que le produit par tête relatif en logarithme, $y_{i,t} = \log(Y_{i,t}) - N^{-1}\sum_i \log(Y_{i,t})$, de l'économie i $(= 1, \dots, N)$ à la date t $(= 1, \dots, T)$ est déterminé par le processus générateur de données suivant,

$$y_{i,t} = (1-\beta)y_{i,t-1} + u_{i,t} \tag{1}$$

$$u_{i,t} \sim iid(0, \sigma_u^2) \text{ dans la coupe et dans la chronologie} \tag{2}$$

$$y_{i,0} \sim iid(0, \sigma_y^2) \text{ et } \mathbb{E}[y_{i,0}u_{j,t}] = 0 \ \forall (i,j,t) \in \{1, \dots, N\}^2 \times \mathbb{N} \tag{3}$$

$$\beta \in [0, 1[\tag{4}$$

Le paramètre β s'interprète comme la vitesse de convergence. Avec ce processus générateur de données, on se limite au cas de la convergence absolue. Cette restriction simplifie l'exposé, plus généralement cela revient à supposer que l'économiste qui adopte l'*approche classique* de la convergence peut parfaitement contrôler l'hétérogénéité des états stationnaires. Il s'agit d'une hypothèse assez peu vraisemblable, si elle n'est pas vérifiée alors les estimateurs envisagés par cette démarche empirique sont éventuellement biaisés (biais de variable omise si celle-ci est corrélée à une des variables explicatives). Ces hypothèses méritent quelques commentaires.

L'hypothèse (1) contraint la vitesse de convergence à être commune dans la coupe des économies. Si on pense au modèle de Solow [140] cela revient à supposer que les économies sont homogènes en termes des paramètres fondamentaux qui déterminent la vitesse de convergence. Il s'agit d'une hypothèse forte, mais on peut difficilement recouvrir l'hétérogénéité éventuelle à partir d'une simple régression en coupe. On verra plus loin comment traiter ce problème en considérant un modèle non linéaire.
L'hypothèse (2) suppose que les chocs idiosyncrasiques affectant les économies sont identiquement et indépendamment distribués. Lever cette hypothèse dans la dimension individuelle n'est pas chose facile, car il faut équiper cette dimension d'une métrique. En évaluant des *distances économiques*, Conley [32] propose une approche non paramétrique pour construire une matrice de variance–covariance et corriger les estimations de la dépendance spatiale. Une application par l'auteur de cette correction non paramétrique

sur une large coupe ne donne pas des résultats sensiblement différents de ceux que nous obtiendrions sans correction de la dépendance spatiale. Mais cette approche repose de façon cruciale sur les variables employées pour mesurer la proximité des économies.
L'hypothèse (3) stipule que la condition initiale du revenu par tête est exogène. Il s'agit d'une hypothèse forte, elle est évidemment liée à l'hypothèse précédente (absence d'autocorrélation temporelle dans la perturbation). Cette hypothèse est très forte car même une version stochastique du modèle de croissance néo-classique, telle que l'on peut la voir dans Den Haan [68], viole cette hypothèse. On verra dans le chapitre [1] que cette hypothèse ne tient plus dès lors que le modèle contient plus d'une variable d'état. Si le modèle contient plus d'une variable d'état, il n'est pas possible de représenter le produit par tête relatif sous la forme d'un processus AR(1).
Enfin, l'hypothèse (4) nous permet d'écarter le cas d'une dynamique explosive.

Une régression en coupe du taux de croissance du produit sur la condition initiale du produit fournit une estimation convergente de la vitesse de convergence et autorise une inférence efficace si et seulement si les hypothèses précédentes sont vérifiées[18]. Les hypothèses d'exogénéité de la condition initiale et de blancheur temporelle de la perturbation sont très fortes. Ceci suggère que cette approche est fragile[19].

L'écriture du processus générateur des données permet de comprendre la critique de Quah reposant sur la *Galton's Fallacy*. En effet, même si les données sont générées par un processus autorégressif d'ordre un asymptotiquement stationnaire ($\beta > 0$), la variance dans la coupe de la condition initiale (σ_y^2) peut être inférieure à la variance de la distribution de long terme du produit relatif. Dans ce cas, la dynamique induit un accroissement régulier

[18]Si la perturbation est caractérisée par une dépendance chronologique, alors on peut obtenir un estimateur convergent si l'asymptotique concerne les dimensions individuelle *et* temporelle. Notons cependant que cette littérature ne développe jamais d'asymptotique dans la dimension temporelle. Si la perturbation est caractérisée par une dépendance spatiale, l'estimateur de la vitesse de convergence est convergent mais les tests sur les paramètres du modèle ne sont pas efficaces.

[19]Voir Evans [49], [50] qui déconseille l'utilisation de cette approche et milite plutôt en faveur d'une approche en séries temporelles.

de la variance en coupe jusqu'à ce qu'elle atteigne son niveau de long terme, alors que les économies convergent au sens de la β-convergence. L'existence d'une force de rappel ne garantit pas que la variance de la distribution du produit se contracte, c'est à dire que les économies se rapprochent. Ainsi, Quah [121] écrit,

> "(...) *a negative correlation between growth rates and levels says nothing about the poor catching up with the rich. Contrary to claims made elsewhere, traditional empirics are completely silent on the important convergence dimension in economic growth.*" Quah [121, page 1048]

On pourrait donc s'interroger sur l'intérêt de la notion de β-convergence. Il faut cependant rendre justice à Barro et Sala-i-Martin. Ces auteurs n'ont jamais affirmé qu'une corrélation négative entre le taux de croissance et le niveau du produit soit en rapport avec un rattrapage des nations pauvres vers les nations riches et ont toujours souligné les différences entre la σ-convergence et la β-convergence. Ces deux concepts nous renseignent sur deux aspects différents de la croissance. La σ-convergence décrit l'évolution de la distribution du produit par tête (en négligeant les moments d'ordre supérieur à deux), la β-convergence s'intéresse à la dynamique à l'intérieur de la distribution du produit par tête. Ainsi, Sala-i-Martin répond aux critiques de Quah de la façon suivante,

> "*This possibility*[20] *led some economists (most prominently Quah (1993)) to criticize the classical approach on three grounds. First they suggested that classical analysts were confusing the two concepts of convergence. Secondly, they argued that the only meaningful concept of convergence was that of σ. Finally, they said that the concept of β-convergence conveyed no interesting information about σ-convergence (or about anything else) so it should not be studied. Needless to say that the three points were not entirely correct. First, classical analysts were well aware of the distinction from the very beginning (...). In fact, that is why they made the distinction in the first place ! Secondly, the intra-distributional mobility (reflected in β) is at least as interesting as the behavior*

[20] Sala-i-Martin se réfère à la possible contradiction entre les deux définitions de la convergence.

of the distribution itself (reflected in σ). In fact, it could even be argued that if mobility was very high, the evolution of σ would be uninteresting. Surprisingly Quah (1994,1996) highlights the importance of intra-distributional mobility in the context of stochastic kernel estimators. Finally, β provides information about σ to the extent that any necessary condition does – the fact that the two phenomena tend to appear together in most data sets seems to support this view." Sala-i-Martin [127, note de bas de page 5]

On peut accorder à Sala-i-Martin, dont le propos ne manque pas d'ironie, que les critiques de Quah sont assez contradictoires. Il faut cependant noter que les démarches considérées par ces deux auteurs pour étudier la mobilité à l'intérieure de la distribution du produit par tête sont radicalement différentes. Cette différence tient au processus générateur des données. Sala-i-Martin, en se référant à la β-convergence, considère une dynamique linéaire alors que Quah ne suppose pas la linéarité de la dynamique.

On peut réconcilier ces deux approches en définissant une β-convergence absolue non linéaire, c'est à dire une spécification non linéaire de la régression en coupe[21] (ou en panel). Cela revient à considérer le processus générateur des données suivant,

$$\Delta y_{i,t} = \mathcal{M}(y_{i,t-1}) + \varepsilon_{i,t} \tag{5}$$

où $\mathcal{M}$ est une fonction continue supposée invariante dans la chronologie. La perturbation satisfait les hypothèses standards (indépendance dans les dimensions individuelle et temporelle). L'hypothèse de β-convergence absolue est vérifiée si la fonction $\mathcal{M}$ est monotone décroissante partout, c'est à dire si on vérifie $\mathcal{M}'(x) < 0$ pour tout $x \in \mathbb{R}$. Ainsi, le taux de croissance espéré d'une économie pauvre est plus important que le taux de croissance espéré d'une économie riche. Ici rien n'empêche que cette condition ne soit vérifiée que localement. Ainsi, le processus des données adopté ici permet de rendre compte d'une multiplicité des régimes de croissance. Les zéros de la dynamique, c'est à dire les valeurs $\mathbf{x}$ telles que $\mathcal{M}(\mathbf{x}) = 0$, définissent les états stationnaires de la dynamique. Le signe de $\mathcal{M}'(\mathbf{x})$ indique la nature stable

[21]Durlauf, Kourtellos et Minkin [46] et Liu et Stengos [92] envisagent aussi une spécification non linéaire de la régression en coupe ou du panel. Mais ils ne font pas explicitement le rapport avec l'approche de Quah.

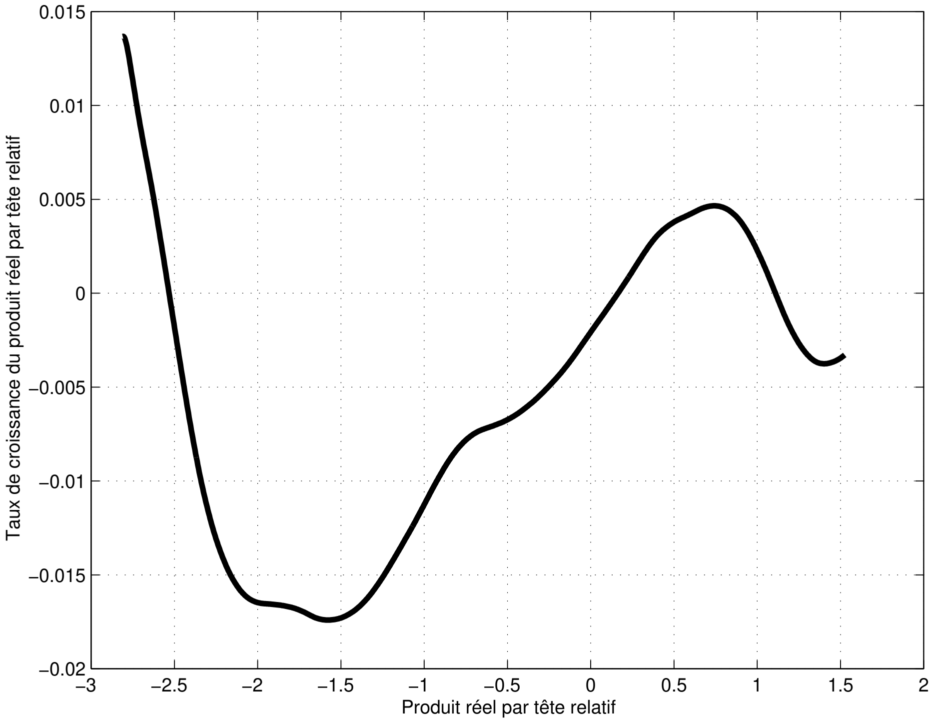

FIG. 5 – β-**convergence absolue non linéaire**.

ou instable de ces états stationnaires. La valeur de $\mathcal{M}'(\mathbf{x})$ indique la vitesse de convergence (ou divergence) locale. L'étude du sens de variation de la fonction $\mathcal{M}(x)$ dévoile les différents régimes de croissance.

La figure [5] représente un estimateur de la régression non linéaire du taux de croissance du produit par tête relatif sur le niveau du produit par tête relatif. On a utilisé le panel dans son entier comme pour l'estimation de la *stochastic kernel*[22]. On compte trois états stationnaires. Les équilibres haut et bas sont stables, tandis que celui du milieu est instable. On discerne ainsi une croissance à régime multiple, de façon cohérente avec les travaux de Quah ou Desdoigts [35]. On distingue deux zones de convergence dans le haut et le bas de la distribution du produit par tête, puis une zone de

[22]On a utilisé un estimateur linéaire locale, voir Pagan et Ullah [106] ou Fan et Gijbels [54] pour une présentation. On peut montrer que le biais de cet estimateur est moins important que celui de Nadaraya-Watson dans les régions où les observations sont moins nombreuses.

divergence dans le milieu de la distribution (si le produit par tête relatif appartient à l'intervalle $[-1,5;0,75]$). On note que si le produit par tête d'une nation est inférieur à 0,2, alors en espérance elle se dirige vers l'équilibre bas de la distribution (puisque son taux de croissance est négatif). On retrouve les conclusions de Quah, le centre de la distribution se creuse pour alimenter les pôles haut et bas de la distribution du produit par tête.

La démarche de Quah est plus générale. En effet, il ne se limite pas à l'étude du premier moment de la dynamique et considère tous les moments de la variable aléatoire (y_t, y_{t+T}). On voit ici qu'en posant plus de structure et en ne considérant que l'espérance conditionnelle non linéaire du taux de croissance du produit par tête relatif, on conclut aussi au caractère *twin peak* de la dynamique. Cet accroissement de structure a l'avantage de nous permettre d'obtenir une information supplémentaire par rapport à la démarche reposant sur l'usage des *stochastic kernel*. On peut évaluer la vitesse de convergence pour les états stationnaires haut et bas. Pour cet échantillon, on obtient une vitesse de convergence de 6% pour l'équilibre bas et une vitesse de convergence de 2% pour l'équilibre haut. La démarche proposée ici, à mi-chemin entre l'*approche classique* et la démarche distributive, donne plus d'information sur la mobilité à l'intérieur de la distribution[23].

De façon assez surprenante, Barro et Sala-i-Martin [16, pages 49-52] envisagent une dynamique non linéaire de la même forme que notre estimation de la fonction $\mathcal{M}$. Ils rationalisent une telle dynamique en faisant référence à un modèle de *Big Push*[24]. Durant la transition d'une économie traditionnelle agricole vers une économie industrialisée et de services, une nation expérimente des rendements croissants, même si les technologies traditionnelle et moderne sont à rendements décroissants. Cela peut aussi expliquer pourquoi la vitesse de convergence est plus importante autour de l'équilibre bas (la trappe à pauvreté). En effet, nous pouvons imaginer que la technologie traditionnelle est moins intensive en capital (rappelons que la vitesse de convergence est une fonction décroissante de la part du capital dans le produit).

[23]Il convient de noter cette démarche empirique repose sur des conditions aussi fortes que l'*approche classique*. L'hypothèse d'une perturbation *iid* dans la coupe et dans la chronologie est toujours nécessaire ici.

[24]Voir Murphy, Shleifer and Vishny [99].

Pourtant, les deux auteurs évacuent ce cas en considérant qu'il n'existe pas de faits empiriques cohérents avec l'idée de rendements décroissants, croissants puis décroissants. Le graphique [5] et les conclusions du chapitre [3], montrent que ce jugement est trop rapide.

Enfin, il convient de noter une nouvelle fois que la généralisation non linéaire de la notion de β-convergence absolue proposée ici n'apporte pas plus d'éléments pour discriminer entre une convergence conditionnelle et convergence locale (*clubs* de convergence). En effet, en conditionnant le produit par tête relatif (par exemple par le taux d'épargne) la relation entre le taux de croissance et le niveau apparaît linéaire et décroissante.

L'*approche classique* et la démarche par les *stochastic kernel* reposent sur des hypothèses très fortes. L'hypothèse d'une perturbation *iid* dans la chronologie semble intenable[25]. La démarche par les séries temporelles est en ce sens plus générale. Celle-ci stipule que deux nations convergent si un choc sur l'écart de produit par tête entre ces deux nations n'a pas un effet permanent. Autrement dit, deux nations convergent s'il existe une force de rappel assurant qu'un choc sur la distance entre ces deux nations ne persiste pas indéfiniment. On voit ainsi que cette définition est proche de la définition de la β-convergence. Il s'agit de tester l'existence d'une force de rappel, *ie* $\beta > 0$ dans l'équation (1), sans poser la blancheur de la perturbation. La différence essentielle est le champs d'application de ces deux démarches empiriques. Comme le note Bernard et Durlauf [21] les régressions en coupes sont plus appropriées si on pense que la dynamique mondiale est une dynamique de transition[26]. En effet, dans ce cas un rapprochement des économies résulte en une relation négative est significative entre la condition initiale et le taux de croissance moyen, alors que ceci est en contradiction avec une définition reposant sur les propriétés des séries temporelles où on suppose que l'hétérogénéité des conditions initiales ne peut être liée à l'hétérogénéité des taux de

[25] Pour que l'usage d'une *stochastic kernel* ait un sens il faut que le produit par tête relatif soit un processus markovien d'ordre un.

[26] Et si on n'adopte pas cette approche pour estimer la vitesse de convergence mais seulement pour tester la β-convergence (*ie* évaluer la corrélation entre le taux de croissance et le niveau).

croissance (qui doit être nulle)[27]. Ceci explique pourquoi l'approche par les séries temporelles donne régulièrement des résultats contre-intuitifs. Même sur des petits groupes de nations relativement homogènes, elle ne parvient pas à rejeter la divergence.

Mesurer la technologie

De la même façon que pour la littérature sur la croissance, la littérature concernée par la quantification du rôle de la technologie dans la croissance ne s'est intéressée à la question de la convergence qu'à partir du milieu des années 80. La technologie est traditionnellement identifiée par la Productivité Globale des Facteurs (PGF), c'est à dire la part de la croissance du produit non expliquée par la croissance des *inputs* (se reporter, par exemple, à Chambers [29, pages 230 à 249]). Il convient de noter, avec Abramovitz [1], que cette mesure de la technologie est résiduelle. Ainsi, la PGF pourrait autant s'interpréter comme une mesure de notre ignorance qu'une mesure du progrès technologique.

En suivant Islam [78] on peut distinguer trois approches pour calculer ou estimer des indices de croissance ou des niveaux de la PGF : (*i*) une démarche exploitant la dimension temporelle (comme, par exemple, Solow [141]), (*ii*) une démarche exploitant la dimension individuelle (comme, par exemple, Hall et Jones [69]) et (*iii*) une démarche exploitant les deux dimensions (comme dans Islam [77]).

Un article représentatif de la première démarche est la contribution de Solow [141]. Celui-ci établit la possibilité de construire des indices de croissance de la PGF sans recourir à des hypothèses paramétriques sur la forme de la fonction de production agrégée. Il se donne une fonction de production à rendements constants et avec un progrès technique Hicks-neutre, $Y(t) =$

[27] Voir la proposition 6 dans Bernard et Durlauf [21, pages 170-171].

$A(t)F(X(t))$, où $X(t)$ désigne un ensemble de facteurs de production à l'instant t. Le processus $A(t)$ exprime les sauts qui affectent la fonction de production, pour toutes les combinaisons d'*inputs*. En examinant la dérivée logarithmique de cette fonction de production, on décrit l'évolution du produit comme des déplacement le long de la fonction de production, via les facteurs de production $X(t)$, et des déplacements de la fonction de production, via le taux de croissance du progrès technique Hicks-neutre. Les productivités marginales, qui déterminent les parts des *inputs* dans le produit, ne sont pas directement observables, c'est pourquoi on se place dans un cadre de concurrence parfaite et d'absence d'externalités, où les prix réels des *inputs* doivent correspondre aux productivités marginales. Ainsi, la donnée des prix réels des facteurs et des quantités, les hypothèses sur la forme du progrès technique, sur la concurrence et l'absence d'externalités, permettent d'obtenir une mesure du progrès technologique.
Si les hypothèses de concurrence parfaite ou d'absence d'externalité paraissent trop fortes, on peut préférer une approche économétrique mais celle-ci demande la spécification d'une fonction de production et de la distribution des erreurs. De plus, elle a le désavantage d'imposer l'indépendance entre les erreurs et les *inputs*.

Hall et Jones [69] mais aussi Easterly et Levine [48] supposent qu'il existe une fonction de production mondiale Cobb Douglas à rendements constants. En posant une part du capital dans le produit de $1/3$ ils dérivent les niveaux et taux de croissance de la PGF[28]. Ces auteurs concluent en soulignant le rôle essentiel de la technologie dans la diversité des niveaux et des dynamiques du produit. Ils obtiennent donc la PGF par une procédure de calibration ; conscients des limites d'une telle approche, Hall et Jones proposent une méthode plus générale dans la version *working paper* de leur article et obtiennent des résultats très proches. Ils affirment qu'il est possible de reprendre la méthode de Solow [141] dans un cadre spatiale. Ils supposent l'existence d'une unique fonction de production mondiale concave à rendements constants. Ainsi le produit de l'économie θ est donnée par $Y(\theta) = A(\theta)F(K(\theta), H(\theta))$, où $K(\theta)$ représente le stock de capital physique, $H(\theta) = h(\theta)L(\theta)$ la force

[28]Hall et Jones comparent des niveaux de la PGF alors que Easterly et Levine comparent des taux de croissance.

de travail pondérée par le niveau d'éducation et $A(\theta)$ la productivité Hicks-neutre. En supposant qu'il existe un continuum ordonné d'économies, ils dérivent alors en suivant Solow [141] une équation de comptabilité de la croissance dans la dimension spatiale,

$$\frac{\dot{y}(\theta)}{y(\theta)} = \frac{\dot{A}(\theta)}{A(\theta)} + \frac{F_1(k(\theta), h(\theta))k(\theta)}{y(\theta)}\frac{\dot{k}(\theta)}{k(\theta)} + \left(1 - \frac{F_1(k(\theta), h(\theta))k(\theta)}{y(\theta)}\right)\frac{\dot{h}(\theta)}{h(\theta)} \quad (6)$$

où k et y sont des variables intensives. Puis ils adoptent une version discrète de l'équation (6),

$$\Delta \log y_i = \bar{\alpha}_i \Delta \log k_i + (1 - \bar{\alpha}_i)\Delta \log h_i + \Delta \log A_i \quad (7)$$

où $\bar{\alpha}_i$ est la part moyenne du capital physique dans le produit pour deux économies consécutives $\bar{\alpha}_i = 0,5(\alpha_i + \alpha_{i-1})$. En normalisant la PGF de l'économie 1 ils obtiennent la PGF de l'économie i de la façon suivante,

$$\log A_i = \sum_{j=2}^{i} \Delta \log A_j + \log A_1$$

Cette démarche ne formule pas d'hypothèse précise sur la forme de la technologie. Elle paraît cependant très discutable. En effet, Hall et Jones supposent que l'on peut ordonner des nations comme on ordonne des années[29]. Le choix de l'ordre des économies a évidemment un impact sur les résultats. De plus, le choix d'un ordre ne se limite pas à quelques alternatives, il existe une infinité de directions possibles. On peut donc difficilement étudier la sensibilité de cette méthode à l'ordre sélectionné. L'interprétation d'un tel exercice semble difficile. L'indice de croissance du A construit par Solow [141] possède une interprétation naturelle. La dynamique du A représente les déplacements de la frontière de production. Ici l'interprétation de la "dynamique" du A est moins évidente, car au départ Hall et Jones posent l'existence d'une fonction de production mondiale commune. Le A mesuré par ces deux auteurs doit correspondre à l'idée d'une distance à une frontière de production commune. Les nations ne peuvent toutes se placer long de la frontière de production

[29] En ce sens, la démarche empirique proposée par Hall et Jones est assez proche de celle proposée par BSM ou MRW. Ces derniers détournent le modèle de Solow [140] vers la dimension individuelle. Hall et Jones détournent Solow [141] vers la dimension individuelle.

commune en exploitant pleinement la technologie mondiale. Ainsi plus le niveau de A_i obtenu par Hall et Jones est important plus la nation i est proche de la frontière de production mondiale. Si cette interprétation est la bonne, la démarche proposée par ces deux auteurs n'est pas très satisfaisante. En effet on aimerait obtenir une estimation de la frontière et pouvoir dire quelles sont les économies capables d'exploiter pleinement la technologie mondiale. On reviendra sur ce point dans le chapitre [3] en proposant l'utilisation d'une méthode plus adaptée dans cette perspective.

Islam [77] propose une démarche exploitant les dimensions individuelle et temporelle pour estimer les niveaux de PGF. Sa démarche est proche de la contribution de Mankiw, Romer et Weil [95]. Ces trois auteurs montrent, en considérant le modèle de Solow [140] à l'état stationnaire, que le produit par tête est déterminé de la façon suivante si la fonction de production est Cobb-Douglas,

$$\log \frac{Y_t}{L_t} = \log A_0 + gt + \frac{\alpha}{1-\alpha} \log(s) - \frac{\alpha}{1-\alpha} \log(n+g+\delta) \tag{8}$$

où A_0 est la PGF initiale, g le taux de croissance commun de la technologie, α la part du capital dans le produit, n le taux de croissance de la population, δ le taux de dépréciation du capital et s le taux d'épargne exogène. L'estimation du terme lié à la PGF est impossible si on utilise, comme MRW, une régression en coupe. Ces trois auteurs supposent que la variabilité dans la coupe de la productivité est non corrélée avec les variables explicatives (taux d'épargne en capital physique et dans une spécification augmentée taux d'épargne en capital humain) en plaçant ce terme dans la perturbation du modèle. Ils proposent alors une estimation du pouvoir explicatif du mécanisme néo-classique dans l'hétérogénéité des niveaux observés. Le R^2 de la régression fournit directement une mesure de la part de l'hétérogénéité des niveaux expliquée par l'accumulation de facteurs. Malheureusement l'interprétation du R^2 repose sur la condition d'orthogonalité entre la productivité et le comportement d'accumulation. Islam propose d'utiliser la dimension panel pour estimer les niveaux, supposés constants, de la PGF[30]. Cette dé-

[30] Il considère une version en panel de l'équation (8), puis la deuxième spécification envisagée par MRW où les nations n'évoluent pas le long de leurs trajectoires de long terme.

marche est en plusieurs points plus séduisante que celle proposée par Hall et Jones [69] ou Easterly et Levine [48]. Premièrement Islam obtient les niveaux de la PGF en utilisant une approche économétrique. En particulier, la part du capital est estimée. Elle n'est pas posée à 1/3 ou déduite des prix des facteurs. Deuxièmement, contrairement à Hall et Jones (première version) les estimations de la PGF ne dépendent pas d'un ordre particulier pour les nations. Outre l'hypothèse d'une fonction de production Cobb-Douglas, la démarche d'Islam est limitée dans au moins une direction importante. Il ne peut véritablement tenir compte d'une dynamique de la productivité. Il suppose que le taux de croissance de la PGF est homogène dans la coupe des nations, cette approche ne peut permettre d'expliquer la diversité des dynamiques de croissance par la diversité des dynamiques de technologie. Il reconnaît l'importance de l'hétérogénéité des niveaux de productivité mais ne va pas jusqu'à abandonner l'hypothèse d'un progrès technique commun.

Ces démarches permettent d'évaluer le rôle de la PGF dans l'hétérogénéité des niveaux et des taux de croissance du produit, mais elles ne sont pas directement liées à la question de la convergence. Depuis le milieu des années 80, on s'interroge aussi sur le rôle de la technologie dans le processus de convergence. Une première façon d'envisager cette question est de s'interroger sur la convergence des productivités globales des facteurs. Wolff [153], Bernard et Jones [22] et [23] s'intéressent directement à la convergence des productivités. Wolff, en utilisant des régressions en coupe, montre l'existence d'un phénomène de rattrapage technologique entre les économies du G7 sur la période 1870-1979, celui-ci étant plus évident après la seconde guerre mondiale. Il met en évidence dans le même temps la convergence du capital physique par tête à partir des années 60. Ainsi la convergence des économies du G7 s'explique par deux mécanismes : une convergence technologique et une convergence néo-classique liée à l'accumulation du capital. Le problème des contributions de ces deux mécanismes à la convergence du produit par travailleur n'est pas examinée par Wolff [153].
Bernard et Jones [22] travaillent sur la base d'un échantillon plus large (14 économies de l'OCDE) à un niveau sectoriel. Ils considèrent une stratégie de test de racine unitaire en panel pour tester la convergence des PGF. Ils montrent que les présomptions en faveur de la convergence de la PGF sont

plus fortes dans les secteurs non manufacturiers.
Une limite de ces articles est qu'ils ne considèrent que des échantillons relativement petits dans la dimension individuelle. Ces échantillons regroupent des économies qui convergent (en termes de produit par tête). De plus, ils n'indiquent pas quel est le mécanisme de convergence le plus important. Sur un échantillon plus large, on aimerait connaître les responsabilités respectives du capital physique et de la technologie dans le phénomène de divergence.
Par ailleurs une part de la littérature utilisant les régressions en coupe pour tester la convergence des nations considèrent le produit initial comme une approximation de la PGF initiale et non pas comme une approximation du stock de capital physique initial. Un exemple est donné par Dowrick et NGuyen [39] qui concluent à la convergence des PGF des économies de l'OCDE[31].

Desdoigts [37] et Dowrick et Rogers [42] proposent des approches plus intéressantes dans notre perspective. Ils augmentent le modèle de MRW [95] d'une dynamique de rattrapage technologique exogène. Les démarches proposées par ces auteurs ont l'avantage (i) d'inclure les deux mécanismes de convergence (technologique et néo-classique) dans le même modèle empirique, (ii) de considérer une large coupe d'économies. Desdoigts [37] utilise, comme MRW, des régressions en coupe alors que Dowrick et Rogers considèrent un panel[32]. Desdoigts [37] adopte l'idée de Nelson et Phelps [102] et pose la dynamique suivante pour la technologie de l'économie θ,

$$\frac{\dot{A}(\theta,t)}{A(\theta,t)} = \Phi(H(\theta)) \log\left(\frac{T(t)}{A(\theta,t)}\right)$$

où $T(t)$ est la meilleure pratique, son taux de croissance g est commun dans la coupe, $A(\theta,t)$ représente le niveau de la technologie pour l'économie θ et $H(\theta)$ sa dotation en capital humain. La fonction Φ est croissante et nulle si $H = 0$. Ainsi le taux de croissance de la technologie est, durant la transition, spécifique à chaque économie. Une économie qui ne parvient pas à mettre en œuvre la meilleure pratique a un taux de croissance technologique supérieur aux taux de croissance de la meilleure pratique (g). Ici la vitesse de l'ajus-

[31]Se reporter à Fagerberg [53] pour une présentation des régressions en coupe dans cette perspective.
[32]On reviendra plus loin, dans le chapitre [1], sur la contribution de ces deux auteurs.

tement technologique, spécifique à chaque économie, est déterminé par la dotation en capital humain. Cette formulation de la dynamique rend compte de l'idée selon laquelle la technologie ne se diffuse pas instantanément et qu'il est d'autant plus facile pour une nation de se rapprocher de la meilleure pratique que sa main d'œuvre est bien éduquée. Desdoigts montre que l'on ne peut rejeter cette spécification augmentée du modèle de MRW contre une spécification sans dynamique hétérogène de la technologie. Il montre ainsi que la prise en compte de cette imparfaite diffusion de la technologie permet de mieux reproduire la distribution du produit par tête en 1995.

Plan de la thèse

Dans cette thèse on s'intéresse essentiellement aux rôles respectifs de la technologie et de l'accumulation de capital physique dans la dynamique de croissance mondiale et dans les phénomènes de convergence ou divergence des régions ou nations.

Dans le chapitre [1], on suppose l'existence de deux mécanismes de convergence : le mécanisme néo-classique (lié à l'accumulation de capital physique) et le mécanisme technologique (lié au rattrapage technologique). En supposant (i) que les paramètres fondamentaux et la vitesse de rattrapage technologique sont communs dans la coupe, (ii) que la vitesse de convergence technologique est inférieure à la vitesse de convergence néo-classique, on cherche à savoir ce qu'une régression du taux de croissance du produit par tête sur son niveau initial peut nous apprendre. Avec l'hypothèse (ii), on n'introduit pas seulement une imparfaite diffusion de la technologie dans le cadre néo-classique. On suppose aussi que cette diffusion est très lente. On verra que la forte persistance supposée des écarts technologiques permet de reproduire des périodes de divergence et de réconcilier (quantitativement) le modèle de croissance avec les faibles estimations de la vitesse de convergence. Ce modèle est caractérisé par deux variables d'état (le capital physique et la technologie), la vitesse de convergence mesurée par une régression en coupe dépend alors de l'horizon temporel T, considéré pour calculer le taux de croissance moyen, et des sources de l'hétérogénéité du produit par tête initial. Si l'hétérogénéité initiale du produit par tête dépend essentiellement de l'hétérogénéité

sur le stock de capital physique initial alors la vitesse de convergence est une fonction décroissante de l'horizon T. Si l'hétérogénéité initiale s'explique essentiellement par une hétérogénéité des niveaux de la technologie, alors la vitesse de convergence adopte un profil croissant en fonction de T. Ainsi, une série de régressions (en augmentant progressivement l'horizon T) mal spécifiées par rapport au processus générateur des données peut nous informer sur les sources de la convergence. On peut donc voir si la convergence résulte d'un *capital deepening effect* ou plutôt d'un *technological catch-up effect*, sans disposer de données sur le capital physique ou la PGF. On a vu plus haut que les régressions du taux de croissance sur le niveau possèdent une double interprétation. Dans certains cas la condition initiale du produit correspond à une *proxy* du stock de capital physique initial (BSM), dans d'autres cas il s'agit d'une *proxy* pour le niveau initial de la technologie (Baumol). La démarche proposée dans le chapitre [1] peut nous permettre de sélectionner une des deux interprétations. Une limite de cette démarche empirique est qu'elle impose l'homogénéité des paramètres fondamentaux[33]. Ainsi il nous paraît raisonnable de l'appliquer seulement sur des données régionales ou sur des petits groupes de nations relativement homogènes. Une application sur les états américains entre 1940 et 2000, montre que la convergence régionale aux États Unis est essentiellement induite par des différences en terme de dotations en capital physique.
Dans le même chapitre on propose une stratégie empirique pour estimer les vitesses de convergence néo-classique et technologique. Cette démarche, contrairement à l'*approche classique*, exploite pleinement la dimension temporelle et repose sur un modèle de croissance stochastique. Une application sur les États Unis se révèle assez décevante mais ne contredit pas notre hypothèse d'une vitesse de convergence technologique inférieure à la vitesse de convergence néo-classique.

Dans le chapitre [2] nous revenons sur les méthodes employées pour tester la convergence des nations. On a vu plus haut que l'approche par les séries temporelles est susceptible d'être la plus robuste. Cette démarche souffre pourtant de quelques défauts. Premièrement, il existe un résultat singulier concluant, de façon contre intuitive, à la convergence des nations sur une

[33]Notons que la littérature ne se préoccupe que très rarement de ce problème.

large coupe. Evans et Karras [52], en utilisant une stratégie de test de racine unitaire en panel, concluent à la convergence sur un échantillon de 54 nations alors que d'autres stratégies (régression en coupe, σ-convergence ou approche distributive) concluraient plutôt à la divergence. Dans le chapitre on montre les limites de la stratégie proposée par ces deux auteurs et on montre que ces limites peuvent expliquer leur étonnant résultat. Il est important d'établir ce point car (i) des stratégies similaires sont régulièrement employées dans la littérature[34], (ii) dans le chapitre suivant nous travaillons sur un échantillon proche de celui-ci.
Un second problème est que sur des échantillons de nations homogènes, comme les états européens ou les régions américaines, l'approche par les séries temporelles ne parvient pas à rejeter la divergence[35] alors que toutes les autres mesures convergent pour conclure à la convergence. Bernard et Durlauf [21] expliquent cette incapacité en notant que cette démarche réclame que les nations évoluent le long de leurs trajectoires de long terme. L'approche par les séries temporelles rejette difficilement la divergence si les données sont caractérisées par une dynamique de transition, même si cette transition résulte en un rapprochement des économies. Pour dépasser cette limite on propose d'associer la transition à la partie déterministe du processus d'écart de produit par tête. On propose une stratégie empirique pour tester cette idée et étudie ses propriétés asymptotiques. En associant la transition à la partie déterministe du processus d'écart de produit par tête on ne parvient pas à rejeter la convergence absolue de dix états européens.

Dans le chapitre [3] on met en œuvre un exercice de comptabilité de la croissance sur une large coupe d'économies. L'objectif de cet exercice est double. En premier lieu, on s'interroge sur les contributions respectives de la PGF et de l'accumulation de capital physique dans la dynamique mondiale de croissance. On montre que le rôle de la PGF est globalement faible par rapport à l'accumulation de capital physique. Mais ces contributions respectives ne sont pas stables dans la coupe et la chronologie. Il apparaît que, si la contribution de la PGF est faible voire négligeable pour les miracles économiques du sud-est asiatique, la dynamique de la PGF est essentielle pour

[34] Voir, par exemple, Gaulier, Hurlin et Jean Pierre [61].
[35] Exception faite de l'approche proposée par Evans et Karras.

rendre compte des désastres économiques. Dans la chronologie, il apparaît que la dynamique de la PGF est globalement marginale entre 1965 et 1990 mais qu'elle est essentielle pour comprendre l'hétérogénéité des taux de croissance du produit par travailleur dans les années 80.

En second lieu, on s'intéresse aux sources de la divergence des nations. Il apparaît que la dynamique du capital physique suffit à reproduire l'aspect *twin peak* de la distribution du produit par travailleur en 1990. La PGF vient renforcer la divergence des nations les plus pauvres.

Pour mettre en œuvre ces exercices de comptabilité de la croissance et de la divergence, on ne peut utiliser les méthodes évoquées plus haut. Celle-ci sont inadaptées pour traiter une large coupe d'économies (l'approche de Solow [141]) ou inadaptées pour autoriser une hétérogénéité des dynamiques de la PGF (l'approche d'Islam [77] en panel). Comme on ne peut se contenter de l'approche par calibration, employée par Easterly et Levine [48], on doit se diriger vers une autre approche. On utilise l'approche *Data Envelopment Analysis*[36] (DEA), pour estimer les indices de croissance de la PGF. Il s'agit d'une approche non paramétrique qui procède en estimant à chaque date une frontière de production mondiale concave en enveloppant les observations d'*inputs-outputs*. En disposant de ces estimations, on peut alors facilement construire des indices de croissance de la PGF. Un des avantages de cette méthode, outre l'aspect non paramétrique, est que cet indice peut se décomposer en un indice reflétant les mouvements de la frontière de production (le progrès technologique) et un indice reflétant l'évolution des distances des nations à la frontière de production mondiale (le rattrapage technologique). Kumar et Russell [84] ont récemment appliqué cette méthode sur le même échantillon que dans le chapitre [3], mais on remet en question leurs conclusions. Ils concluent comme nous au faible rôle globale de la PGF dans la croissance, mais aussi à l'existence d'un rattrapage technologique (les économies se rapprochent de la frontière de production mondiale en moyenne). On montre que ce dernier résultat ne tient pas si le processus générateur des données interdit les régressions technologiques et si on adopte une stratégie empirique cohérente avec cette hypothèse.

Dans le même chapitre on utilise la méthode DEA et une généralisation de

[36]Se reporter à Färe, Grosskopf, Norris et Zhang [56] pour une première application sur données macro-économiques.

celle-ci pour s'interroger sur une hypothèse généralement adoptée dans la littérature empirique de la croissance et de la convergence. On propose un test de l'hypothèse de concavité de la fonction de production mondiale. Ce test est intéressant à plusieurs titres. En premier lieu il nous permet de vérifier si les approches empiriques généralement employées ne placent pas des hypothèses trop fortes sur les données. Par exemple, on suppose souvent que la fonction de production de production mondiale est Cobb-Douglas à rendements constants (voir MRW [95], Benhabib et Spiegel [18],...). En second lieu le test de l'hypothèse de concavité de la fonction de production mondiale peut apporter un élément intéressant dans le débat convergence locale – convergence conditionnelle. En effet, on sait que si la convexité de l'ensemble de production n'est pas vérifié, alors rien n'assure l'unicité de l'état stationnaire. Dans ce cas, même si les économies sont homogènes, le niveau de long terme de chaque nation dépend de sa condition initiale. L'application sur une très large coupe d'économies en 1995 conclut à la non concavité de la fonction de production mondiale.

Chapitre 1

L'*approche classique*

1 Introduction

Un fait régulièrement observé, sur des données internationales ou régionales, est l'instabilité chronologique de la vitesse de convergence estimée par une régression à la Barro. Ce chapitre propose une explication de ce phénomène et propose d'en tirer parti pour ré-interpréter les régressions à la Barro.

Barro et Sala-i-Martin [15], étudient la convergence des états américains et montrent que la vitesse de convergence, estimée sur des sous périodes de 10 ans, est instable. La deuxième colonne du tableau [1.1] reporte les estimations de BSM [15]. Les deux auteurs utilisent une statistique de ratio de vraisemblance pour tester la stabilité chronologique de la vitesse de convergence et concluent que celle-ci n'est pas constante. La figure [1.1] représente des estimations de la vitesse de convergence pour 48 états américains[1] en glissement décennales entre 1929 et 2000. La vitesse de convergence est clairement instable. Si on considère les intervalles de confiance au seuil de 5%, la vitesse de convergence est comprise entre 14% et -2%. Elle apparaît souvent non significativement différente de zéro à partir de 1970. La figure [1.2] propose une représentation alternative de l'instabilité de la vitesse de convergence. Dans ce graphique, on fixe la condition initiale en 1940 et on fait varier

[1]On ne considère pas l'Alaska et Hawaii car les données ne sont disponibles qu'à partir de 1969 pour ces deux états.

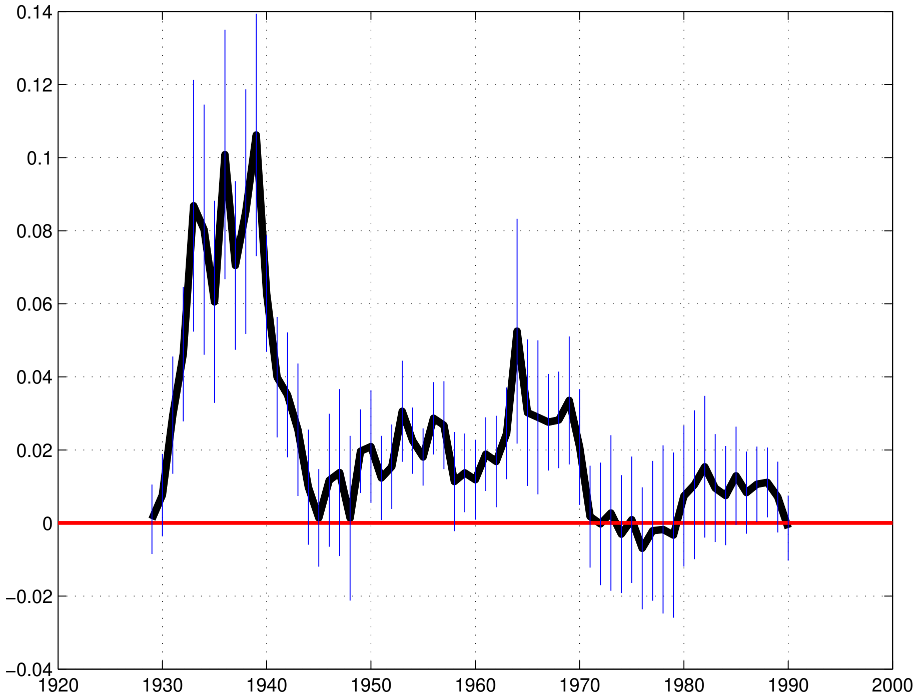

FIG. 1.1 – **Vitesses de convergence pour 48 états américains (I).** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête net des transferts sur la condition initiale et des indicatrices régionales. Les barres verticales représentent des intervalles de confiance au seuil de 5%. Les vitesses de convergence sont obtenues sur des périodes de dix ans.

l'horizon temporel de la régression en coupe. Encore une fois la vitesse de convergence apparaît instable.

Le graphique [1.3, en page 101] représente des estimations de la vitesse de convergence pour 3075 comtés américains, inclus dans les 48 états considérés plus haut, en glissement décennales entre 1969 et 2000. Cette fois-ci, la vitesse de convergence est toujours significativement positive. Si on considère les intervalles de confiance au seuil de 5%, la vitesse de convergence est comprise entre 5% et 0,5%. La vitesse de convergence est instable et on discerne, dans les deux graphiques un profil globalement décroissant de la vitesse de convergence. Ce profil décroissant de la vitesse de convergence est observé par ailleurs. Par exemple, Martin [97] montre que la vitesse de convergence

Fig. 1.2 – **Vitesses de convergence pour 48 états américains (II).** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête net des transferts sur la condition initiale et des indicatrices régionales. Les barres verticales représentent des intervalles de confiance au seuil de 5%. Les vitesses de convergence sont obtenues en augmentant l'horizon temporel de la régression en coupe.

entre les régions européennes est tombée de 2% à 1,3% dans les années 80. Le graphique [1.4, en page 102] montre l'évolution de la vitesse de convergence pour les économies de l'OCDE. Cette fois, le profil de la vitesse de convergence n'est pas aussi clair.

BSM [15] proposent une explication de l'instabilité de l'estimateur de la vitesse de convergence. Il est vraisemblable que les chocs affectant les économies soient corrélés dans la dimension individuelle. Sous certaines conditions, ceci peut résulter en un biais d'endogénéité qui brouille la mesure de la vitesse de convergence. On sait que certains riches états américains (comme le Texas) sont producteurs de pétrole. Une augmentation temporaire du prix du pétrole tend à augmenter le taux de croissance de ces états et ainsi à sous estimer la vitesse de convergence. Plus formellement ils posent une structure à facteur commun pour les chocs affectant les états. Le modèle considéré par les deux auteurs est de la forme suivante[2],

$$g_{i,t,t-1} = a - (1 - e^{-\beta})[\log y_{i,t-1} - x(t-1)] + u_{i,t} \tag{1.1}$$

$$u_{i,t} = \phi_i s_t + v_{i,t} \tag{1.2}$$

$g_{i,t,t-1}$ représente le taux de croissance du produit par tête de l'état i entre t et $t-1$, $y_{i,t-1}$ le niveau en $t-1$ du produit par tête de l'état i, x est le taux de croissance exogène de la technologie neutre au sens de Harrod et β la vitesse de convergence. Le facteur commun s_t s'interprète comme un choc agrégé, la variable aléatoire ϕ_i indique comment ce choc agrégé affecte les différents états, enfin, la variable aléatoire $v_{i,t}$ est supposée indépendamment et identiquement distribué dans les dimensions temporelle et spatiale. Si ϕ_i est corrélé avec la condition initiale, c'est à dire s'il y a une relation entre la condition initiale d'un état et sa réaction face à un choc agrégé, alors la condition initiale est corrélée à la perturbation $u_{i,t}$, et l'estimation de la vitesse de convergence est biaisée. Ce modèle est assez simple pour que l'on puisse déterminer le sens du biais. Si la corrélation entre la condition initiale et ϕ_i est positive, alors la vitesse de convergence est sous estimée.

BSM proposent de contrôler ce type d'effet en augmentant le modèle estimé d'une variable rendant compte de la composition du produit dans chaque état. Cette variable a pour but de mimer le terme $\phi_i s_t$ dans (1.2). La troisième colonne du tableau [1.1] reporte leurs estimations de la vitesse de convergence lorsqu'ils contrôlent de la composition du produit. La statistique du ratio de vraisemblance montre que cette fois-ci on accepte l'homogénéité chronologique de la vitesse de convergence au seuil de 5%. Si BSM parviennent ainsi à expliquer l'instabilité de la vitesse de convergence estimée, on aimerait comprendre pourquoi la vitesse de convergence adopte un profil décroissant.

[2]BSM [15, pages 228-229]. Cette formulation est employée dans les travaux récents qui cherchent à rendre compte de la dépendance spatiale dans les panels non stationnaires, voir, par exemple, Phillips et Sul [116] et le chapitre [2, en page 119].

Lorsque le modèle considéré comporte une unique variable d'état, la vitesse de convergence ne doit pas dépendre du temps dans un voisinage de l'état stationnaire. Si la condition initiale est trop éloignée de l'état stationnaire, dans le sens où l'approximation autour de l'état stationnaire n'est pas valable, la vitesse de convergence estimée n'est plus constante dans la chronologie. Les estimations reportées par BSM pourraient bien s'expliquer de cette façon. Barro et Sala-i-Martin [16, page 84] simulent numériquement la vitesse de convergence en fonction de la distance à l'état stationnaire, dans un modèle à une variable d'état (le capital physique), sans utiliser une approximation linéaire. La vitesse de convergence diminue à mesure que l'économie se rapproche de l'état stationnaire. Dans la suite on cherche une autre explication de l'instabilité de la vitesse de convergence mesurée par une régression à la Barro. Deux raisons motivent ce choix. Premièrement, expliquer la variabilité de la vitesse de convergence par une distance initiale trop importante à l'état stationnaire est une réponse externe et on préfère rester dans le cadre standard en supposant que cette approximation est valable. Deuxièmement, cette explication ne peut produire un profil croissant de la vitesse de convergence ni produire une période de divergence des économies (comme pour les économies de l'OCDE).

Dans ce chapitre on propose une explication alternative. La dynamique, dans le modèle considéré par BSM, est entièrement déterminée par l'accumulation du capital physique. Un tel modèle prédit une vitesse de convergence constante dans un voisinage de l'état stationnaire. On considère une dynamique de croissance plus riche, en augmentant le modèle d'une dynamique de productivité exogène. Ceci permet d'ajouter un mécanisme complémentaire de convergence. L'hétérogénéité du produit par tête s'explique alors par l'hétérogénéité sur la productivité et/ou l'hétérogénéité sur le stock de capital par tête. Même dans un voisinage de l'état stationnaire, la vitesse de convergence des économies, mesurée par une régression à la Barro, n'est plus constante. En effet, dans un tel modèle, la dynamique du produit est intimement liée aux causes de l'hétérogénéité caractérisant le produit par tête initial.

En fait, si les données sont générées par ce modèle augmenté les régressions à la Barro sont mal spécifiées, car ces régressions ne sont valables que dans le cadre d'un modèle à une seule variable d'état. Cette spécification inadéquate induit un biais dans la mesure de la vitesse de convergence[3]. Dans la section [2] on donne une expression analytique de ce biais asymptotique (dans la dimension individuelle) et on montre que la vitesse de convergence obtenue par une régression à la Barro ne peut identifier la vitesse de convergence associée au mécanisme néo-classique. En effet, dans ce modèle augmenté, deux valeurs propres stables sont associées aux deux variables d'état (le capital physique et la productivité), la vitesse de convergence mesurée par une régression à la Barro est un mélange, pas nécessairement convexe, de ces deux valeurs propres.

Néanmoins, les régressions à la Barro peuvent constituer un outil utile dans le débat sur les sources de la croissance et de la convergence. En effet, on peut exprimer le biais asymptotique de l'estimateur de la vitesse de convergence en fonction des sources de l'hétérogénéité du produit par tête initial et de l'horizon chronologique considéré. Ainsi le profil de la vitesse de convergence estimée pour différents horizons peut donner une idée sur les sources de la croissance. On montre en particulier que, si la vitesse de convergence néo-classique est supérieure à la vitesse de convergence technologique, on obtient un profil décroissant de la vitesse de convergence estimée lorsque l'hétérogénéité sur le capital initial est la source essentielle de l'hétérogénéité du produit initial et un profil croissant lorsque l'hétérogénéité sur la productivité initiale est la source essentielle de de l'hétérogénéité du produit initial. Dans le premier cas c'est le *capital deepening effect* qui l'emporte dans l'explication de la dynamique de convergence. Dans le second cas c'est le *technological catch-up effect* qui explique la dynamique de convergence.

Si les données sont générées par ce modèle, on ne peut estimer les deux vitesses de convergence en utilisant une régression à la Barro augmentée. Cette approche consisterait à régresser le taux de croissance du produit par tête sur une constante et les conditions initiales des deux variables d'état (le stock de capital physique et la productivité). Même si on dispose des don-

[3]Le biais asymptotique est défini comme l'écart entre la limite en probabilité (dans la dimension individuelle) de l'estimateur de la vitesse de convergence et l'opposé de la plus grande valeur propre stable de la dynamique.

nées nécessaires, les vitesses de convergence ne sont pas identifiables si on considère un modèle où l'épargne est déterminée de façon endogène[4]. Dans la section [3] on propose une approche empirique alternative pour estimer les deux vitesses de convergence. On considère un modèle de croissance stochastique où les chocs technologiques se diffusent imparfaitement. Si la tendance de progrès technique Harrod-neutre est de la forme $A_{t+1} = A_t(1+g)e^{x_{t+1}}$ avec $x_{t+1} = \rho x_t + \varepsilon_{t+1}$ et $\varepsilon_t \sim \; iid(0, \sigma_\varepsilon^2)$, alors on montre que le taux de croissance du produit par travailleur est un processus ARMA(2,2). Le paramètre $1 - \rho$ s'interprète comme la vitesse de convergence technologique. Plus ρ est important, plus la diffusion de l'innovation ε_{t+1} est lente. On montre que les racines de la partie de la partie autorégressive correspondent aux vitesses de convergence technologique et néo-classique (liée à l'accumulation du capital physique). On applique cette approche sur les États Unis, les résultats ne sont pas totalement satisfaisants.

La suite du chapitre se découpe donc de la façon suivante. La section [2] envisage le cas d'un modèle néo-classique de croissance augmenté d'une dynamique de *catch-up* technologique exogène. Elle propose une ré-interprétation des régressions à la Barro et offre des évidences empiriques, sur données régionales et internationales. La section [3] présente une version stochastique du modèle présenté dans la section [2], cela permet de proposer deux stratégies d'estimation des vitesses de convergence néo-classique et technologique. Les conclusions sont spécifiques à chaque section. Les annexes techniques, tableaux et graphiques sont renvoyés dans les sections suivantes.

[4]Supposer un comportement d'épargne exogène reviendrait à admettre l'orthogonalité entre la dynamique d'accumulation du capital et la productivité.

2 Convergence & Technologie

2.1 Introduction

Cette section caractérise l'estimateur de la vitesse de convergence associé à une régression à la Barro, lorsque les données sont générées par un modèle comprenant plus d'une variable d'état. On se donne un modèle de croissance optimale néo-classique augmenté d'une dynamique de rattrapage technologique exogène. Ce modèle offre deux explications complémentaires du phénomène de convergence des économies : le mécanisme néo-classique – lié à l'accumulation du capital physique – et le mécanisme technologique.

Den Haan [68] a déjà souligné que dans un tel modèle la dynamique du produit est étroitement liée à l'explication de la déviation initiale du produit par tête à sa cible. Si la déviation initiale est essentiellement expliquée par la productivité (c'est à dire par un choc qui affecte directement le produit et donc affecte a priori la consommation et l'investissement dans le même sens) le modèle induit un chemin de retour vers l'état stationnaire très différent du cas où la déviation initiale est liée à une destruction du capital (les réactions initiales de la consommation et de l'investissement sont opposées). Dans le dernier cas on a un retour monotone vers l'état stationnaire alors que dans le premier cas le retour est non monotone (on observe une phase de divergence) et plus lent. Pour Den Haan [68], ceci permet d'expliquer les estimations faibles de la vitesse de convergence obtenues sur de nombreux échantillons. A l'aide de simulations par Monte Carlo, il montre que la vitesse de convergence obtenue avec une régression à la Barro dépend de l'explication de la condition initiale sur le produit par tête. Ces simulations lui permettent de conclure que si, parmi les états Américains, les déviations initiales du produit par tête sont liées à la productivité alors une part du capital de un tiers est cohérente avec une vitesse de convergence de l'ordre de 2%.

Cette section offre une approche analytique en caractérisant ce qu'une régression à la Barro permet d'identifier dans un tel modèle asymptotiquement (*ie* lorsque dimension individuelle tend vers l'infini). Cette approche complète en plusieurs points les résultats de Den Haan [68]. En supposant que la vitesse de convergence néo-classique est supérieure à la vitesse de convergence technologique[5], on établit les résultats suivants. ***(i)*** On encadre la limite en probabilité de l'estimateur de la vitesse de convergence. En particulier on montre que celle-ci est toujours inférieure à la vitesse de convergence néo-classique. ***(ii)*** On dérive les conditions nécessaires et suffisantes assurant que la limite en probabilité de l'estimateur de la vitesse de convergence est comprise entre la vitesse de convergence néo-classique et la vitesse de convergence technologique. ***(iii)*** On montre que la limite en probabilité de l'estimateur de la vitesse de convergence est une fonction de l'horizon T sur lequel est considéré le taux de croissance moyen. Ceci permet d'expliquer le profil décroissant, régulièrement estimé sur des échantillons divers, de la vitesse de convergence. Lorsque T tend vers l'infini, la limite en probabilité de la vitesse de convergence tend vers la vitesse de convergence technologique. ***(iv)*** Comme suggéré par les simulations de Den Haan, la limite en probabilité de l'estimateur de la vitesse de convergence dépend de façon croissante du ratio des variances du capital et de la productivité (en écarts à l'état stationnaire). Des estimations sur des données régionales et internationales suggèrent que l'estimation régulièrement reportée[6] d'une vitesse de convergence autour de 2% est plus liée à la persistence des chocs de productivité qu'à la dynamique néo-classique associée aux rendements décroissants du capital.

La section [2.2] présente le modèle considéré dans cette section. La section [2.3] propose une étude analytique des régressions à la Barro. La section [2.4] propose une application aux états et comtés des États Unis. La section [2.5] conclut. Les annexes sont regroupées La section [4, en page 87] regroupe les annexes, graphiques et tableaux associés à la section [2].

2.2 Un modèle néo-classique de rattrapage

La technologie de production est de type néo-classique, le produit, $Y(t)$, est obtenu à partir du capital, $K(t)$ et du travail efficace $\mathcal{T}(t)L(t)$: $Y(t) = A(t)F(K(t), \mathcal{T}(t)L(t))$. $\mathcal{T}(t)$ représente le progrès technologique neutre au sens de Harrod, on pose $\mathcal{T}(t) = e^{xt}$ avec $x > 0$ le taux de croissance technologique exogène. $L(t)$ représente le nombre de travailleurs à l'instant t, on pose $L(t) = e^{nt}$ avec $n > 0$ le taux de croissance exogène de la force de travail.

[5]Den Haan adopte implicitement cette hypothèse.

[6]Voir Sala-i-Martin [127].

Le terme $A(t) \in]0,1]$ représente la productivité, assimilée ici à la distance à la frontière technologique définie par $F(\bullet,\bullet)$. Ce terme de productivité est caractérisé par une dynamique de rattrapage exogène, *ie* $\dot{A} = \phi(1 - A(t))$. Le paramètre $\phi > 0$ défini la vitesse de rattrapage technologique, une plus grande valeur de ce paramètre se traduit par une réduction du temps que mettra l'économie pour combler son retard technologique. Pour caractériser la technologie de production on pose $\alpha(\hat{k}(t))$ et $\varepsilon(\hat{k}(t))$ respectivement la part du capital dans la valeur ajoutée et l'élasticité entre capital et travail, avec $\hat{k}(t) \equiv \frac{K(t)}{\mathcal{T}(t)L(t)}$ (voir l'annexe pour une définition formelle de ces paramètres).

L'utilité intertemporelle est définie par,

$$\int_0^\infty e^{-(\rho-n)t} u(c(t))dt$$

où $u(\bullet)$ est une fonction d'utilité instantanée vérifiant les propriétés habituelles et $c(t)$ et la consommation par travailleur (ou tête). Pour caractériser les préférences on pose $\sigma(t)$ l'inverse de l'élasticité de substitution intertemporelle.

Le programme du planificateur social associé à ce modèle est défini par,

$$\begin{aligned}
&\underset{\{c(t),t\in\mathbb{R}\}}{\mathcal{M}ax} \int_0^\infty e^{-(\rho-n)t} u(c(t))dt \\
&\qquad \underline{sc}\ \dot{k}(t) = A(t)F(k(t),\mathcal{T}(t)) - (\delta+n)k(t) - c(t) \\
&\qquad \dot{A}(t) = \phi(1 - A(t)) \\
&\qquad k(0),\ A(0) \text{ donnés.}
\end{aligned}$$

où $k(t)$ et δ sont respectivement le stock de capital par travailleur à l'instant t et le taux de dépréciation du capital. Ce modèle est caractérisé par deux variables d'état ($A(t)$ et $k(t)$) et une variable de contrôle ($c(t)$). La dynamique linéarisée autour de l'état stationnaire de cette économie doit donc être caractérisée par deux valeurs propres négatives et une valeur propre explosive. Les deux valeurs propres négatives sont liées à la vitesse de convergence néoclassique μ et la vitesse de convergence technologique ϕ. Comme le processus de convergence technologique est indépendant du processus de convergence néo-classique, il est assez naturel de trouver,

$$\mu = -\frac{\xi}{2} + \frac{1}{2}\sqrt{\xi^2 + \frac{4(1-\alpha)}{\sigma\varepsilon}(\rho + \sigma x + \delta)\left[\frac{\rho + \sigma x + \delta}{\alpha} - \delta - x - n\right]}$$

où $\xi \equiv \rho - n - (1-\sigma)x$.

La vitesse de convergence néo-classique (μ) dépend en particulier de l'élasticité de substitution entre le capital et le travail (ε). On voit directement que la vitesse de convergence est une fonction décroissante du paramètre ε. Le sens de cette relation est intuitif. En effet, à mesure de l'élasticité de substitution entre les facteurs, lorsque l'économie s'ajuste vers sa cible elle échange du travail contre du capital. On dira que la croissance est d'autant plus riche en capital que l'élasticité de substitution entre capital et travail est importante. Face à un choc négatif sur le niveau de capital, la réaction initiale de l'investissement est d'autant plus faible que ε est fort car les ménages anticipent que la croissance va s'enrichir en capital (*ie* une baisse de la productivité marginale future du capital) et préfèrent limiter la baisse de la consommation face au choc[7]. Ce mécanisme contribue à ralentir l'ajustement et explique la relation décroissante entre μ et ε. Évidemment cet effet ne peut s'exercer indépendamment de l'élasticité de substitution intertemporelle. On note que ε est multiplié par σ dans l'expression de la vitesse de convergence néo-classique. Ce terme croisé signifie que le mécanisme identifié plus haut joue moins si les ménages ont une forte élasticité de substitution intertemporelle (*ie* une forte volonté de lissage, σ faible). On verra plus loin que l'élasticité de substitution entre les facteurs est non négligeable dans l'identification des sources de la croissance.

Autour de l'état stationnaire, la dynamique des variables d'état est décrite

[7]Une explication peut être plus convaincante tient plus directement à la forme de la fonction de production. L'élasticité de substitution entre les facteurs est liée à la courbure de la fonction de production. Plus ε est important, plus la fonction de production est linéaire (on se rapproche de la parfaite substitution entre les facteurs). Une destruction de capital augmente d'autant moins la productivité marginale du capital que la fonction de production est linéaire. Ainsi, l'incitation des ménages à investir contemporainement au choc est d'autant plus faible que ε est fort, car la productivité marginale est plus faible.

par,

$$\tilde{A}(t) = e^{-\phi t}\tilde{A}(0) \tag{1.3}$$

$$\tilde{k}(t) = e^{-\mu t}\tilde{k}(0) + (e^{-\mu t} - e^{-\phi t})\frac{\varepsilon}{1-\alpha}\gamma^{-1}\tilde{A}(0) \tag{1.4}$$

où $\tilde{A}(t) \equiv \log A(t)$ et $\tilde{k}(t) = \log \hat{k}(t) - \log \hat{k}^*$ représentent respectivement l'écart à l'état stationnaire de la productivité et du stock de capital par unité de travail efficace, et avec,

$$\gamma = \frac{\sigma\varepsilon}{1-\alpha}\frac{(\xi+\phi)\phi - \frac{1-\alpha}{\sigma\varepsilon}(\rho+\sigma x+\delta)\left(\frac{\rho+\sigma x+\delta}{\alpha} - \delta - n - x\right)}{(\rho+\sigma x+\delta)\left(\frac{\rho+\sigma(x+\phi)+\delta}{\alpha} - \delta - n - x\right)} \tag{1.5}$$

On peut montrer (voir l'annexe, section [4.1, en page 87]) que γ est négatif si et seulement si la vitesse de convergence technologique est inférieure à la vitesse de convergence néo-classique. Ainsi, le signe du terme devant la productivité en déviation à l'état stationnaire, dans l'équation (1.4), est positif pour toutes les valeurs des paramètres fondamentaux. Une inefficacité technologique initiale d'une économie a toujours un impact négatif sur le niveau du capital. On note aussi que cet effet est d'autant plus fort que la dynamique de croissance est riche en capital. L'argument est analogue à l'explication de la relation décroissante entre μ et ε. Face à un choc de productivité négatif, l'investissement et la consommation chutent. La chute de l'investissement est plus forte si l'élasticité de substitution entre capital et travail est forte car les ménages anticipent un enrichissement de la croissance en capital et donc une productivité marginale future du capital plus faible. Cette chute plus importante de l'investissement explique pourquoi l'effet d'un choc de productivité sur le capital est plus important lorsque ε est fort. Encore une fois, on peut voir que ce mécanisme n'est pas indépendant de l'élasticité de substitution intertemporelle. L'effet d'un choc de productivité sur le produit dépend des préférences des ménages, via la dynamique d'accumulation du capital physique.

On déduit la dynamique du produit dans un voisinage de l'état stationnaire à partir de la dynamique des variables d'état,

$$\tilde{y}(T) = \alpha e^{-\mu T}\tilde{k}(0) + (e^{-\mu T} - e^{-\phi T})\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\tilde{A}(0) + e^{-\phi T}\tilde{A}(0) \tag{1.6}$$

et le taux de croissance moyen du produit par tête

$$g_T = x + \frac{1-e^{-\mu T}}{T}\alpha \log \hat{k}^* - \frac{1-e^{-\mu T}}{T}\alpha \log \hat{k}(0) - \frac{1-e^{-\phi T}}{T}\log A(0) + \frac{e^{-\mu T}-e^{-\phi T}}{T}\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\log A(0) \tag{1.7}$$

avec $g_T \equiv T^{-1}(\log y(T) - \log y(0))$ le taux de croissance moyen du produit par travailleur. On voit alors que l'on ne peut définir l'ajustement du produit à sa cible sans s'interroger sur ce qui explique l'écart initial entre le produit et sa cible de long terme. Selon que cette distance s'explique par un niveau de capital trop faible ou par une incapacité à exploiter efficacement la technologie le produit par tête exhibe des trajectoires très différentes. Den Haan [68] a déjà souligné ce point, la réaction du produit dépend de la nature du choc. Le graphique [1.5, en page 103] montre que les réponses du produit peuvent être très différentes alors même que l'écart initial du produit à sa cible est identique. Les cas 1,4,2 et 3 correspondent respectivement aux cas où $(\tilde{k}(0), \tilde{A}(0)) = \{(-0.3, -0.1); (-0.6, 0); (0.3, -0.3); (0, -0.2)\}$. Le cas 4 correspond au modèle sans dynamique de la productivité, dans les autres cas le modèle est augmenté. Dans le cas 4 l'écart initial du produit est expliqué par une destruction de capital. Celle-ci induit contemporainement une augmentation de l'investissement (car la productivité marginale a augmentée), une baisse de la consommation et du produit. A mesure que le capital se rapproche de sa cible, l'incitation à investir diminue et la consommation comme le produit rejoignent leurs cibles respectives de façon monotone. Si la déviation initiale du produit à sa cible est essentiellement expliquée par un choc de productivité (c'est à dire un choc qui affecte directement le produit), alors contemporainement la consommation et l'investissement chutent. Ceci induit une appréciation de la productivité marginale du capital entraînant donc une augmentation de l'investissement et une baisse de la consommation (et du produit). L'augmentation de l'investissement finit par porter ses fruits, la productivité marginale du capital augmente puis le profil de la consommation (et du produit) se retourne. Dans ce cas le retour du produit à sa cible n'est pas monotone. On observe une phase de divergence du produit qui peut être assez longue.

Dans une perspective de comparaison internationale, ces simulations suggèrent que les économies peuvent adopter des profils très différents en termes de convergence vers l'état stationnaire. Même si les paramètres fondamentaux sont homogènes, on a une hétérogénéité des dynamiques de convergence dans la coupe. Dès lors on peut légitimement s'interroger sur ce qu'une régression à la Barro peut nous apprendre à propos la dynamique mondiale du produit.

2.3 Régression à la Barro

Dans cette section on veut évaluer ce qu'une régression à la Barro[8] permet d'identifier si les donnés sont conformes au modèle décrit dans la section précédente. Afin de simplifier l'exposé, on suppose que les économies sont homogènes en termes des paramètres fondamentaux et on se concentre sur la convergence absolue. L'hétérogénéité des économies provient seulement des conditions initiales sur le stock de capital par tête et la productivité. On suppose de plus que le monde est formé d'un continuum d'économies. De façon plus formelle le processus générateur des données s'écrit,

$$g_T(\theta) = x - \frac{1-e^{-\mu T}}{T}\alpha\theta_1 - \frac{1-e^{-\phi T}}{T}\theta_2 + \frac{e^{-\mu T}-e^{-\phi T}}{T}\frac{\alpha\epsilon}{1-\alpha}\gamma^{-1}\theta_2 + u(\theta) \quad (1.8)$$

où $\theta_1 = \tilde{k}(0)$ et $\theta_2 = \tilde{A}(0)$ représentent respectivement le capital par tête en déviation à l'état stationnaire et la productivité en déviations à l'état stationnaire. On note $\theta \equiv (\theta_1, \theta_2)'$. Le monde est essentiellement caractérisé par la distribution de θ. On note $\varphi(\theta)$ la densité associée à cette distribution, Φ sa fonction de répartition et Θ le support de cette variable aléatoire bivariée. Le terme $u(\theta)$ représente une perturbation d'espérance nulle vérifiant $\int_\Theta \theta u(\theta) \mathrm{d}\Phi(\theta) = 0$. Il peut s'agir d'une erreur de mesure sur le taux de croissance du produit ou de variables omises qui expliquent la croissance mais qui sont orthogonales aux conditions initiales. Le modèle estimé est de la forme suivante,

$$g_T(\theta) = a - \frac{1-e^{-bT}}{T}\log y(\theta, 0) + v(\theta)$$

où $g_T(\theta)$ est le taux de croissance moyen de l'économie θ, $\log y(\theta, 0)$ est le produit par tête initiale de l'économie θ, $v(\theta)$ une perturbation. On estime

[8] Voir Barro et Sala-i-Martin [16, sections 2.6.5 et 11.2.1]

ce modèle par les moindres carrés non linéaires en cherchant les valeurs de a et b telles que le critère suivant est minimisé,

$$\int_\Theta \left(g_T(\theta) - a + \frac{1-e^{-bT}}{T} \log y(\theta, 0) \right)^2 \mathrm{d}\Phi(\theta) \tag{1.9}$$

Les conditions nécessaires (et suffisantes) sont données par les deux équations suivantes,

$$\int_\Theta \left(g_T(\theta) - \hat{a}_\infty(T) + \frac{1-e^{-\hat{b}_\infty(T)T}}{T} \log y(\theta, 0) \right) \mathrm{d}\Phi(\theta) = 0 \tag{1.10}$$

$$\int_\Theta \log y(\theta,0) \left(g_T(\theta) - \hat{a}_\infty(T) + \frac{1-e^{-\hat{b}_\infty(T)T}}{T} \log y(\theta, 0) \right) \mathrm{d}\Phi(\theta) = 0 \tag{1.11}$$

Minimiser le critère (1.9) revient à mettre en œuvre des moindres carrés non linéaires sur un échantillon de dimension infinie. Ainsi on dérive directement des résultats asymptotiques en minimisant le critère (1.9). Ce que l'on étudie ici c'est la limite en probabilité (lorsque le nombre d'individus dans la coupe tend vers l'infini) de l'estimateur de la vitesse de convergence issu d'une régression à la Barro.

Les propositions suivantes cherchent à caractériser le comportement asymptotique de l'estimateur de b déterminé par ces deux équations normales non linéaires. La première proposition suppose que les conditions initiales sur le capital physique et la productivité ne sont pas corrélées. La deuxième proposition établit que cette hypothèse ne change pas qualitativement les résultats.

Proposition 2.1. *Supposons que $\mathbb{C}ov[\theta_1, \theta_2] = 0$ et que l'on ait $\phi < \mu$. A distance finie ($T < \infty$) la vitesse de convergence obtenue à l'aide d'une régression à la Barro est donnée par l'expression suivante,*

$$\hat{b}_\infty(T) = \phi - \frac{1}{T} \log B(T)$$

avec

$$\frac{\mathbb{V}[\tilde{y}(\theta,0)]}{\mathbb{V}[\theta_2]}B(T) = e^{-(\mu-\phi)T}\left(\alpha^2\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} + \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right) + \left(1 - \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)$$

(i) *$\hat{b}_\infty(T) \to \phi$ lorsque $T \to \infty$.* ***(ii)*** *On peut encadrer la vitesse de convergence obtenue par une régression à la Barro $\underline{b}(T) \leq \hat{b}_\infty(T) \leq \mu$, avec*

$$\underline{b}(T) = \phi - \frac{1}{T}\log\left\{1 + \left(e^{-(\mu-\phi)T} - 1\right)\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right\}$$

On a $\phi < \hat{b}_\infty(T) < \mu$ si et seulement si l'inégalité suivante est vérifiée,

$$\frac{\alpha^2\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} > -\frac{\alpha\varepsilon}{(1-\alpha)}\gamma^{-1} \tag{1.12}$$

(iii) *La vitesse de convergence mesurée par une régression à la Barro est une fonction monotone décroissante, à partir d'un rang T^*, de la dimension chronologique si et seulement si l'inégalité (1.12) est vérifiée (il y a au plus un changement de variation).*

Preuve Voir la section [4.3, en page 94]. ■

On obtient des résultats analogues si $\phi > \mu$ (*ie* $\hat{b}_\infty(T) \to \mu$), de façon plus générale, la vitesse de convergence issue d'une régression à la Barro tend vers l'opposé de la plus grande valeur propre stable de la dynamique lorsque l'horizon T tend vers l'infini. Cette proposition montre, de façon explicite, ce qu'une régression à la Barro permet d'identifier lorsque le modèle est augmenté d'une variable d'état. Il apparaît clairement que la vitesse de convergence ainsi mesurée est sans rapport direct avec le mécanisme de convergence néo-classique. Dès lors les discussions sur la cohérence d'une faible vitesse de convergence et la part du capital dans la valeur ajoutée semble vaine.
Notons que la vitesse de convergence mesurée par une régression à la Barro ($\hat{b}_\infty(T)$) n'est pas nécessairement comprise entre les vitesses de convergence néo-classique et technologique.

On voit qu'un déterminant essentiel de la vitesse de convergence est le ratio de la variance de θ_1 à la variance de θ_2. A distance finie ($T < \infty$), la

limite en probabilité de la vitesse de convergence estimée par une régression à la Barro dépend de l'état initial du monde.
$\hat{b}_\infty(T)$ tend[9] vers la vitesse de convergence technologique ϕ par dessus si et seulement si l'hétérogénéité sur le capital par tête est une source assez importante de l'hétérogénéité de la condition initiale du produit par tête. Si la fonction de production est Cobb-Douglas (*ie* $\varepsilon = 1$), l'hétérogénéité de la condition initiale du produit par tête s'explique en majorité par l'hétérogénéité sur le stock de capital physique initial si et seulement si $\alpha^2\mathbb{V}[\theta_1] > \mathbb{V}[\theta_2]$. Ainsi, on aimerait que le seuil à partir duquel on observe $\hat{b}_\infty(T) \searrow \phi$ lorsque $T \to \infty$ soit $\alpha^2\mathbb{V}[\theta_1]/\mathbb{V}[\theta_2] > 1$. Mais la condition (1.12) est moins sévère[10]. Ce résultat semble intuitif, le membre de droite de l'inégalité (1.12) est inférieur à l'unité car la dynamique du capital n'est pas indépendante de la productivité. Même si θ_1 n'est pas corrélé avec θ_2, le modèle induit une corrélation entre $\tilde{k}(\theta, T)$ et $\tilde{A}(\theta, T)$ (pour $T < \infty$). En reprenant les solutions du modèle données dans la section précédente, on peut montrer que cette covariance est donnée par,

$$\mathbb{C}ov(\alpha\tilde{k}(\theta,T), \tilde{A}(\theta,T)) = \left\{e^{-(\mu-\phi)T} - 1\right\} \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\mathbb{V}[\tilde{A}(\theta,T)]$$

Cette covariance est positive pour toutes les valeurs admissibles des paramètres fondamentaux. A une date $T < \infty$, le modèle induit une corrélation positive expliquée par l'action de la productivité sur la dynamique capital[11]. Ainsi, pour que la dynamique mondiale soit dominée par la déviation initiale du capital par tête à son état stationnaire, au sens où la vitesse de convergence mesurée est supérieure à ϕ, il n'est pas nécessaire d'avoir $\alpha^2\mathbb{V}[\theta_1] > \mathbb{V}[\theta_2]$. On voit facilement qu'il existe une valeur seuil ε^* telle que le membre de droite de l'inégalité (1.12) est inférieure à l'unité si et seulement si $\varepsilon < \varepsilon^*$. Pour la calibration retenue du modèle[12] on a $\varepsilon^* = 1,272$.

---

[9] Lorsque l'horizon T tend vers l'infini.

[10] Pour la calibration retenue (et en imposant $\varepsilon = 1$), le membre de droite de l'inégalité (1.12) est inférieur à l'unité. Ce résultat demeure valide dans un voisinage de la calibration retenue (toujours avec $\varepsilon = 1$).

[11] La nullité asymptotique de la covariance, comme des autres moments du second ordre, est liée à la nature déterministe du modèle. Il faudrait perturber le modèle à chaque date par des petits chocs de productivité pour que les moments d'ordre deux ne tendent pas vers zéro, voir la section [3, en page 72].

[12] Se reporter aux tableaux [1.2, en page 112] et [1.3].

Même si une régression à la Barro est mal spécifiée dans ce cadre, elle peut apporter quelques informations au débat sur les sources de la croissance. Sans explicitement mettre en œuvre un exercice de comptabilité de la croissance (car, par exemple, on n'observe pas le capital et la productivité), on peut avoir une idée des contributions respectives du capital et de la productivité. Si la vitesse de convergence adopte un profil décroissant en fonction de T, c'est que la dynamique du capital est plus importante que la dynamique de la productivité pour expliquer la croissance mondiale. En adoptant ce critère, on ne tient pas compte de la corrélation entre le capital par tête et la productivité induite par le modèle. Cela conduit, d'une certaine façon, à sous évaluer le rôle de la technologie en omettant de considérer l'effet induit de la productivité sur la dynamique du capital. Les exercices de comptabilité de la croissance sont soumis à la même limite (voir le chapitre [3, en page 164]).

La proposition [2.1] montre que la vitesse de convergence mesurée par une régression à la Barro est toujours inférieure à μ et peut être inférieure à ϕ. On comprend facilement pourquoi cette vitesse de convergence peut être inférieure à ϕ. En effet, si l'hétérogénéité initiale s'explique seulement par l'incapacité des nations à exploiter efficacement la technologie mondiale ($\mathbb{V}(\theta_2) > 0$ et $\mathbb{V}(\theta_1) = 0$) on sait que cette inefficacité va entraîner une baisse du stock de capital physique, c'est à dire induire une seconde source d'hétérogénéité du produit. Ceci résulte en un allongement de l'ajustement du produit. Si l'hétérogénéité initiale des nations s'explique seulement par les dotations en capital physique ($\mathbb{V}(\theta_2) = 0$ et $\mathbb{V}(\theta_1) > 0$), la vitesse de convergence mesurée par une régression à la Barro correspond nécessairement à μ, pour tout T, car dans ce cas il n'y a pas de dynamique de la productivité (rappelons que celle-ci est supposée exogène). Dans un cas intermédiaire, la vitesse de convergence mesurée par une régression à la Barro est nécessairement inférieure à μ car l'enrichissement de la dynamique du produit par une dynamique de productivité ne peut que ralentir l'ajustement du produit vers sa cible. La vitesse de convergence mesurée par une régression à la Barro ne peut donc être supérieure à la vitesse de convergence néo-classique (μ).

Le graphique [1.6, en page 104] représente la limite en probabilité de la vitesse de convergence estimée par une régression à la Barro pour différentes valeurs du ratio $\mathbb{V}[\theta_1]/\mathbb{V}[\theta_2]$ et différents horizons T[13]. Il apparaît que si ce ratio est unitaire et si l'horizon est inférieur à dix ans, alors la vitesse de convergence mesurée est négative. On obtient un seuil, en deçà duquel la vitesse de convergence mesurée est négative, plus important si la dynamique de productivité est plus persistante (*ie*, ϕ plus petit) ou si la variance de θ_2 est plus grande relativement à celle de θ_1. Un modèle de croissance néo-classique peut générer une relativement longue phase de divergence des économies. Une augmentation de σ, c'est à dire une baisse de l'élasticité de substitution intertemporelle conduit aussi à une augmentation du seuil. En effet, si cette élasticité de substitution est plus faible alors la chute initiale de la consommation est plus petite relativement à celle de l'investissement (les ménages préfèrent consommer le plus possible contemporainement au choc). L'incitation à investir, et donc à baisser la consommation, par la suite est plus grande. La consommation en déviation à l'état stationnaire (comme le produit) va baisser durant une période plus longue pour aider à la reconstitution du stock de capital par travailleur efficace. De la même façon, une augmentation de l'élasticité de substitution entre capital et travail induit une hausse du seuil.

Dans la proposition 2.1 nous avons supposé l'absence de corrélation entre la distribution du stock de capital par tête initial en déviation à l'état stationnaire et la distribution initiale de la productivité. La proposition suivante envisage le cas où la corrélation est non nulle.

Proposition 2.2. *Supposons que la vitesse de convergence néo-classique soit supérieure à la vitesse de convergence technologique ($\phi < \mu$) et que $\mathbb{V}[\theta_2] + \mathbb{C}ov(\alpha\theta_1, \theta_2) > 0$[14]. On note $\chi \equiv \mathbb{C}ov(\alpha\theta_1, \theta_2)$. A distance finie ($T < \infty$) la vitesse de convergence obtenue à l'aide d'une régression à la Barro est donnée par l'expression suivante,*

$$\hat{b}^c_\infty(T) = \phi - \frac{1}{T}\log B^c(T)$$

[13]Notons que, pour la calibration retenue, on a $T^* = 0$.

[14]Cette condition assure que la vitesse de convergence soit définie pour tout T.

avec

$$\frac{\mathbb{V}[\tilde{y}(\theta,0)]}{\mathbb{V}[\theta_2]+\chi}B^c(T) = e^{-(\mu-\phi)T}\left(\frac{\alpha^2\mathbb{V}[\theta_1]+\chi}{\mathbb{V}[\theta_2]+\chi}+\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)+\left(1-\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)$$

On retrouve tous les résultats de la proposition 2.1 en remplaçant la condition (1.12) par l'inégalité suivante,

$$\frac{\alpha^2\mathbb{V}[\theta_1]+\chi}{\mathbb{V}[\theta_2]+\chi} > -\frac{\alpha\varepsilon\gamma^{-1}}{(1-\alpha)} \tag{1.13}$$

Preuve Voir la section [4.4, en page 97]. ■

La prise en compte de la covariance entre θ_1 et θ_2 ne change pas fondamentalement le message donné dans la proposition 2.1. La vitesse de convergence mesurée à l'aide d'une régression à la Barro tend vers l'opposé de la plus grande valeur propre stable de la dynamique. Cette vitesse de convergence est donc toujours sans rapport direct avec le mécanisme de convergence néoclassique. La vitesse de convergence ainsi mesurée dépend toujours, à distance finie, de l'état initial du monde.

On peut réécrire la condition (1.13) de la façon suivante,

$$\alpha^2\mathbb{V}[\theta_1]+\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\mathbb{V}[\theta_2] > -\left(1+\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)\mathbb{C}ov(\alpha\theta_1,\theta_2)$$

Supposons que l'on ait $\varepsilon = 1$, c'est à dire que la technologie soit Cobb-Douglas. Si la covariance entre θ_1 et θ_2 est négative alors cette condition est plus restrictive que la condition (1.12). En effet, pour la calibration retenue, avec $\varepsilon = 1$, le terme précédant la covariance vaut -0,233[15]. Il faut choisir une variance de θ_1 plus forte relativement à celle de θ_2 pour assurer que la dynamique du capital domine la dynamique du produit, au sens où $\hat{b}^c_\infty(T) > \phi$. Ce résultat est intuitif. Dans le cas d'une covariance négative une économie initialement incapable d'exploiter efficacement la frontière technologique mondiale ($A(\theta,0) < 1$) est aussi (en espérance) dans une situation de suraccumulation du capital. Cette économie n'est pas incitée à investir, deux

[15]Ce terme demeure négatif si on perturbe les paramètres fondamentaux (sauf ε).

motifs de baisse de l'investissement coexistent. Si la covariance est positive, il y a un motif pour baisser l'investissement initial ($A(\theta, 0) < 1$) et un motif pour augmenter l'investissement initial (la sous-accumulation du capital). Cette contradiction contribue à diminuer le temps d'ajustement des économies. Ainsi, il n'est pas nécessaire d'avoir une forte hétérogénéité initiale du capital pour augmenter la vitesse de convergence mesurée par une régression à la Barro en la rapprochant de la vitesse de convergence néo-classique.
On peut à nouveau voir que la sévérité de la condition (1.13), lorsque la covariance est négative, dépend de l'élasticité de substitution entre capital et travail. Si cette dernière est faible, le terme devant la covariance devient plus important en valeur absolue, la condition est alors plus restrictive que la condition (1.13) (lorsque la covariance est négative).

Les graphiques [1.7, en page 105] et [1.8, en page 106] représentent la vitesse de convergence mesurée par une régression à la Barro dans le cas où les ratios $\mathbb{V}[\theta_1]/\mathbb{V}[\theta_2]$ sont respectivement 1 et 20, en faisant varier la covariance de façon à ce que la corrélation balaye l'intervalle $[-1, 1]$. Quelle que soit la covariance, si l'hétérogénéité du capital est de même grandeur que celle de la productivité, la vitesse de convergence mesurée par une régression à la Barro est inférieure à ϕ pour tout T.

Les propositions précédentes montrent comment on peut inférer l'importance de l'effet de *capital deepening* par rapport au mécanisme de *catch-up* technologique, à partir d'une série de régressions à la Barro. La proposition suivante montre qu'il est possible d'avoir une idée du ratio des variances du capital initial et de la productivité initiale à partir d'une série de régression à la Barro si on ne dispose pas d'*a priori* sur la corrélation entre les conditions initiales.

Proposition 2.3 (Partage des eaux). *Supposons que la vitesse de convergence néo-classique soit supérieure à la vitesse de rattrapage technologique. Il existe une fonction,*

$$b^*(T) = \phi - \frac{1}{T} \log \frac{e^{-(\mu-\phi)T} + 1 + \left(e^{-(\mu-\phi)T} - 1\right) \frac{\alpha\varepsilon}{1-\alpha} \gamma^{-1}}{2} \tag{1.14}$$

telle que $\alpha^2\mathbb{V}[\theta_1] > \mathbb{V}[\theta_2]$ si et seulement si $\hat{b}^c_\infty(T) > b^*(T)$. La fonction $b^*(T)$ est une fonction décroissante de l'élasticité de substitution entre capital et travail, si $\varepsilon = \varepsilon^*$ on a $b^*(T) = \phi$.

Preuve Voir la section [4.5, en page 98]. ■

Quelle que soit la corrélation entre θ_1 et θ_2, si on mesure une vitesse de convergence à l'aide d'une régression à la Barro au dessus de $b^*(T)$ alors on a nécessairement $\alpha^2\mathbb{V}[\theta_1] > \mathbb{V}[\theta_2]$. Pour la calibration retenue, si la vitesse de convergence mesurée appartient à la surface grisée du graphique [1.10, en page 108] alors on sait que $\alpha^2\mathbb{V}[\theta_1] < \mathbb{V}[\theta_2]$. On retrouve une nouvelle fois la valeur seuil (ε^*) et rôle crucial pour l'élasticité de substitution entre les facteurs. Pour un ratio $\alpha^2\mathbb{V}[\theta_1]/\mathbb{V}[\theta_2]$ et un horizon T donnés on obtient une vitesse de convergence mesurée par une régression à la Barro d'autant plus faible que la dynamique de croissance est riche en capital (*ie* que ε est fort). Ainsi la frontière associée à $\alpha^2\mathbb{V}[\theta_1] \lesseqgtr \mathbb{V}[\theta_2]$ se déplace vers le sud lorsque l'élasticité de substitution entre les facteurs augmente.

La proposition [2.3] donne finalement assez peu de précision sur la grandeur du ratio des deux variances. On pourrait réduire l'incertitude relative à la grandeur du ratio lorsque la covariance n'est pas connue en se donnant une croyance (*ie* une distribution) sur la corrélation entre θ_1 et θ_2. Par exemple, on a vu que le modèle implique une corrélation positive entre $\tilde{k}(\theta, T)$ et $\widetilde{A}(\theta, T)$, il apparaît ainsi légitime (en imaginant que le modèle existe avant l'instant 0) de supposer que la corrélation entre θ_1 et θ_0 est positive.

Corollaire 2.4. *On suppose que la vitesse de convergence néo-classique est supérieure à la vitesse de rattrapage technologique et que la corrélation entre θ_1 et θ_2 est positive ou nulle. On suppose de plus que la vitesse de convergence au sens de Barro, notée β, est connue. Notons $\nu = \mathbb{V}[\alpha\theta_1]/\mathbb{V}[\theta_2]$ le ratio d'intérêt. Alors, pour chaque T, ν est encadré par ν^* et ν^{**}, définis de la* *façon suivante,*

$$\nu^* = \frac{e^{-(\mu-\phi)T}\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1} - e^{-(\beta-\phi)T} + 1 - \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}}{e^{-(\beta-\phi)T} - e^{-(\mu-\phi)T}} \tag{1.15}$$

$$\nu^{**} = \sqrt{\frac{-(1-\nu^*) + \sqrt{(1-\nu^*)^2 + 4\alpha^2}}{2\alpha^2}} \tag{1.16}$$

On a $\nu^{**} = 1$ si et seulement si $\nu^* = 1$, ou de façon équivalente si et seulement si $\beta = b^*(T)$.

Preuve Voir la section [4.6, en page 99]. ■

Les propositions [2.1] et [2.2] nous donnent des éléments pour ré-interpréter les régressions à la Barro. Si celles-ci ne peuvent nous donner une information sur les deux vitesses de convergences, elles peuvent nous donner, si on croît au modèle et si on dispose d'*a priori* assez précis sur les paramètres fondamentaux, une idée sur les sources de la convergence. Ces propositions suggèrent que si l'on observe un profil décroissant de la vitesse de convergence au sens de Barro, alors l'incitation à accumuler le capital physique est le moteur le plus important de la croissance. La littérature empirique offre de nombreux cas, où l'on observe une diminution de la vitesse de convergence. Par exemple Barro et Sala-i-Martin [15], montrent que sur la période d'après guerre la vitesse de convergence des états américains adopte un profil décroissant sur la période d'après guerre[16]. Les régressions mises en œuvre par ces deux auteurs ne ressemblent pas aux régressions envisagées dans les propositions [2.1], [2.2] et [2.3], où l'on accroît l'horizon T en gardant la même condition initiale (alors que Barro et Sala-i-Martin envisagent des régressions sur des périodes de dix ans successives). La proposition suivante envisage un cas plus proche des régressions mises en œuvre par BSM. On note $g_T(\theta) = \frac{1}{\tau}\left(\log y(\theta, T+\tau) - \log y(\theta, T)\right)$ le taux de croissance moyen entre T et $T+\tau$, et $\hat{\beta}_\infty(T)$ la limite en probabilité de la vitesse de convergence au sens de Barro entre T et $T+\tau$.

[16]La décroissance n'est pas monotone, mais il faut rappeler que les résultats des propositions sont asymptotiques.

Proposition 2.5. *On peut encadrer la vitesse de convergence $\hat{\beta}_\infty(0)$ de la même façon que dans les propositions [2.1] et [2.2]. Si la vitesse de convergence néo-classique est supérieure à la vitesse de convergence technologique, alors $\hat{\beta}_\infty(T)$ tend vers ϕ.*

Preuve : Suivre la même démarche que pour les propositions [2.1] et [2.2], voir la section [4]. ■

Cette proposition est moins riche que les propositions précédentes. La différence essentielle tient au fait qu'ici on considère une condition initiale différente à chaque instant. Puisque la limite en probabilité de la vitesse de convergence mesurée par une régression à la Barro dépend du ratio des variances du capital initial et de la productivité initiale, la dynamique de la vitesse de convergence adopte un profil plus complexe pour tendre vers ϕ. La figure [1.11, en page 109] montre la trajectoire de la vitesse de convergence lorsque la covariance initiale est nulle et que la variance du capital initial est 10 fois plus importante que la variance de la productivité initiale. On voit que le chemin emprunté par la vitesse de convergence à la Barro pour rejoindre ϕ est moins directe. En particulier la vitesse de convergence passe sous ϕ et rejoint sa cible par dessous. Une telle trajectoire est impossible dans le cadre envisagé par les propositions [2.1] et [2.2].

2.4 Un nouveau regard sur les données

Dans la figure [1.2, en page 43] on s'intéresse à la convergence de 48 états américains entre 1940 et 2000. Le graphique représente les vitesses de convergence estimées obtenues en augmentant graduellement l'horizon de la régression en coupe. Clairement, le profil est décroissant, la vitesse de convergence estimée commence autour de 20% (entre 1940 et 1941) et tend vers 2% (entre 1940 et 2000). Conditionnellement au modèle envisagé plus haut, cela veut dire que le processus de convergence des états américains sur cette période est essentiellement lié à un phénomène de *capital-deepening*. Le retard technologique éventuel de certains états joue moins que les différences en termes de dotations en capital pour expliquer la convergence des états américains. De plus, puisque les vitesses de convergence estimées doivent être bornées supérieurement par la vitesse de convergence néo-classique, on peut voir que celle-ci est au moins supérieure à 20%[17]. Dès lors il apparaît inutile de poser une part du capital physique de l'ordre de 0,75 pour rendre compatible la vitesse de convergence traditionnellement observée et le modèle théorique considéré pour justifier les régressions en coupe. On sait d'après le modèle présenté plus haut que la vitesse de convergence estimée doit tendre vers la vitesse de convergence technologique lorsque la taille de l'échantillon tend vers l'infini. Ici on observe que la vitesse de convergence tend vers 2%. Cela nous permet de donner une nouvelle interprétation à cette vitesse de convergence régulièrement reportée dans de nombreuses études[18]. Ces études qui obtiennent régulièrement une vitesse de convergence de 2% n'identifient pas la vitesse de convergence néo-classique (liée à l'accumulation du capital) mais la vitesse de convergence technologique si l'horizon T est assez grand. Cette régularité empirique ne trouve pas son origine, comme l'avancent certains, dans la robustesse du modèle de convergence conditionnelle, mais s'explique plutôt par la persistance des chocs de productivités. Ainsi le modèle néo-classique ne prédit pas une vitesse de convergence trop importante, et la vitesse de convergence estimée ne remet pas en cause quantitativement le modèle de croissance néo-classique[19].

Dans la figure [1.12, en page 110] on s'intéresse à la convergence de 3075 comtés américains. Une nouvelle fois il apparaît que la source principale de convergence est liée au mécanisme néo-classique. On note que la vitesse de convergence estimée tend vers une vitesse de convergence technologique nettement inférieure à 2%. Dans cet échantillon, la dynamique de productivité est plus persistante.

La figure [1.13, en page 111] aborde le cas des pays de l'OCDE. Dans ce dernier cas le partage entre les deux histoires n'est pas aussi clair, mais il semble que l'histoire technologique soit plus pertinente pour comprendre la

[17]Notons néanmoins que l'encadrement de la vitesse de convergence est un résultat asymptotique.

[18]Voir, par exemple, Sala-i-Martin [127].

[19]"(...) *This capital share is larger than the traditional* $\alpha = 0.3$ *estimated under the assumptions of perfectly competitive economies with no externalities and no human capital. A value of 0.75 suggests that, even though the neoclassical model is qualitatively consistent with the data, from a quantitative point of view, it tends to predict too high a speed of conditional convergence.*", Sala-i-Martin [127, pages 1034-1035].

dynamique de convergence d'après seconde guerre mondiale. Dans ce cas la vitesse de convergence estimée semble tendre, par dessous, vers une vitesse de convergence technologique légèrement supérieure à 2%.

2.5 Conclusion

Le détournement de la régression à la Barro nous permet de regarder les données sous un nouveau jour. Si cette approche ne peut nous permettre d'identifier la vitesse de convergence néo-classique car il est difficile de penser que la dynamique du capital puisse suffire à représenter la dynamique du produit, elle donne néanmoins plusieurs informations intéressantes. Premièrement, le modèle utilisé ici permet d'expliquer pourquoi le phénomène de convergence est instable, c'est à dire pourquoi on estime régulièrement des vitesses de convergence qui varient dans la chronologie. Deuxièmement, cette approche peut indiquer quelles sont les sources de la convergence. On ne propose pas une comptabilité quantitative de la convergence, mais on peut dire si la croissance s'explique essentiellement par le *capital deepening* ou par un retard technologique. Par exemple, sur les données américaines il apparaît clairement que l'effet néo-classique domine le rattrapage technologique. Les faits sont moins tranchés pour les pays de l'OCDE. Troisièmement, notre approche permet de ré-interpréter la régularité empirique, assez miraculeuse, d'une vitesse de convergence de 2%. Den haan [68] a déjà montré qu'une régression en coupe ne peut identifier la vitesse de convergence néo-classique. Dans cette section on a montré que cette approche peut identifier la vitesse de convergence technologique lorsque l'horizon T tend vers l'infini. Ainsi, la régularité empirique, qui a en partie déterminé le succès de l'*approche classique*, s'explique plus par le fait que les dynamiques de productivités soient généralement très persistantes.

On a supposé que la vitesse de convergence technologique (exogène) est inférieure à la vitesse de convergence néo-classique. Ce choix, qui nous permet d'identifier la vitesse de convergence technologique est motivé par plusieurs arguments. Premièrement, ce choix permet au modèle de générer une phase de divergence des nations assez longue et donc d'être cohérent avec avec ce que l'on peut observer. Deuxièmement, cela permet de réconcilier le modèle néo-classique de croissance avec une vitesse de convergence de 2% sans remettre en cause l'hypothèse de concurrence parfaite ou l'hypothèse d'absence d'externalité.

On aimerait pouvoir estimer les vitesses de convergence néo-classique et technologique. Pour cela, si on désire rester dans le cadre de l'*approche classique*, il faudrait être capable d'estimer le modèle suivant,

$$\begin{aligned} g_T = x + \frac{1-e^{-\mu T}}{T}\alpha \log \hat{k}^* \quad &- \quad \frac{1-e^{-\mu T}}{T}\alpha \log \hat{k}(0) - \frac{1-e^{-\phi T}}{T}\log A(0) \\ &+ \quad \frac{e^{-\mu T}-e^{-\phi T}}{T}\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\log A(0) \end{aligned}$$

c'est à dire régresser le taux de croissance du produit par travailleur sur les deux variables d'état du modèle (le capital et la productivité). Même si nous disposions de mesures pour ces deux variables[20] les vitesses de convergence ne sont pas identifiables. Pour rendre ce modèle identifiable il faudrait retirer le dernier terme lié à la dépendance de la dynamique du capital au niveau initial de la technologie et estimer par ailleurs le paramètre α. Cette solution ne pourrait fournir une estimation convergente des vitesses de convergence que si ce dernier terme est effectivement nul (biais d'endogénéité). Ce terme est nul si et seulement si les deux vitesses de convergence sont identiques. Mais dans ce cas on revient alors à la spécification et au modèle de Barro et Sala-i-Martin [15], et on ne peut expliquer l'instabilité chronologique de la vitesse de convergence. Il suffit de régresser le taux de croissance sur le produit initial pour obtenir une estimation convergente de la vitesse de convergence. Une approche alternative serait de supposer que les données ne sont pas générées par un modèle d'épargne optimale. Ainsi l'épargne n'est pas affectée par la dynamique de productivité durant la transition et le dernier terme n'apparaît pas dans le modèle à estimer. Dowrick et Rogers [42] proposent un modèle empirique proche, dans l'esprit, de celui utilisé par MRW. Mais ce modèle est augmenté d'une dynamique exogène de rattrapage technologique. Ils supposent une fonction de production Cobb-Douglas et un comportement d'épargne exogène. Leurs estimations en panel (par GMM) montrent que généralement la vitesse de convergence néo-classique est supérieure à la vi-

[20]Nous pourrions obtenir la variable $\log A(0)$ avec les méthodes non paramétriques présentées dans le chapitre [3].

tesse de convergence technologique (sauf pour un sous échantillons de 22 pays de l'OCDE). Puisqu'ils obtiennent une vitesse de convergence technologique significativement positive, ils concluent que les économies initialement éloignées de la meilleure technologie bénéficient de ce retard[21]. Leur approche est en plusieurs points critiquable. Par exemple, dans leur modèle estimé ils remplacent l'état de la technologie, spécifique à chaque économie, par l'état du produit par travailleur. Si l'utilisation du produit par travailleur comme une *proxy* de la technologie est fréquente[22], son usage peut poser problème. En effet, la variabilité de cette proxy s'explique en partie par l'hétérogénéité des niveaux de capital physique. Ainsi, le produit réel par travailleur contient plus d'information que celle liée au niveau de la technologie. Dès lors l'interprétation de la mesure de la vitesse de convergence technologique proposée par Dowrick et Rogers devient difficile. Ces deux auteurs utilisent cette *proxy* en affirmant que le niveau de la technologie est inobservable. Cette assertion est partiellement fausse dans la mesure où ils disposent de données sur le stock de capital et où ils ont supposé une fonction de production Cobb-Douglas. Ainsi, il devrait être possible d'estimer la vitesse de convergence technologique sans utiliser cette approximation.

Le contenu de cette section nous permet, une nouvelle fois, d'insister sur un problème récurrent dans les travaux empiriques s'inscrivant dans l'*approche classique*. A la suite de Barro [12] de nombreux articles proposent des régressions du taux de croissance du produit sur de nombreuses variables susceptibles d'expliquer la croissance[23]. Mais ces exercices sont rarement théoriquement fondés. Par exemple, on s'interroge rarement sur le statut des variables utilisées (s'agit-il de variables introduites pour contrôler de l'hétérogénéité des états stationnaires ou de variables d'état ?). En l'ab-

[21]On verra dans le chapitre [3], en adoptant une approche plus générale, que cette conclusion positive relativement à l'existence d'un phénomène de rattrapage technologique est contestable. Ce résultat pourrait bien être lié aux hypothèses très fortes adoptées par ces deux auteurs.

[22]Desdoigts [37] utilise la même *proxy*. Il s'intéresse aux contributions respectives de l'accumulation de facteurs et de la technologie à la convergence en utilisant un cadre plus proche de MRW [95] (en particulier, le modèle empirique est une régression en coupe et non pas un panel). Il ne cherche pas, comme Dowrick et Rogers [42], à estimer les deux vitesses de convergence.

[23]Voir Durlauf et Quah [47] pour un recensement récent des variables utilisées dans ces régressions.

sence de modèle économique on ne peut donner d'interprétation au coefficient associé au niveau initial du produit par tête, ni même proposer une interprétation causale des régressions. L'objet de ce genre d'exercice est purement descriptif, ce qui n'est pas sans intérêt. Il s'agit alors d'une sorte d'exercice de *data mining*, dont le but ultime serait la composition d'un dictionnaire des variables corrélées avec le taux de croissance du produit[24].

Pour terminer, notons que le modèle considéré dans cette section est sûrement beaucoup trop simple. Un modèle quelconque avec plus d'une variable d'état sera toujours capable de générer une dynamique de la vitesse de convergence mesurée par une régression à la Barro. Notre choix est déterminé par la simplicité et par l'objet de cette thèse (*ie* le partage entre accumulation de facteurs et technologie comme moteur de la croissance et de la convergence). Évidemment nous pourrions construire une histoire très différente pour expliquer l'instabilité chronologique du processus de convergence. Le modèle devrait être prolongé dans au moins deux directions.

Premièrement nous avons supposé une dynamique de productivité exogène. La dynamique de l'efficience est déterminée par un processus autorégressif d'ordre un, la productivité retourne de façon monotone vers son niveau de long terme ($A^* = 1$). On aimerait pouvoir reproduire une dynamique non monotone de l'efficience en l'endogénéisant. En effet, nous verrons dans le chapitre [3] que certaines économies rattrapent la meilleure technologie alors que d'autres s'en éloignent. Il serait donc souhaitable que le modèle considéré soit capable de générer une divergence technologique temporaire (à l'instar de la dynamique du produit par travailleur).

Deuxièmement, nous avons travaillé dans le cas d'économies autarciques. Cette approximation pourrait sembler beaucoup trop forte lorsque l'on travaille, par exemple, sur la convergence des états américains. Dans le cas d'économies parfaitement intégrées le mécanisme de transition néo-classique

[24]Ce travail n'est pas des plus simple, surtout si l'on cherche à classer par ordre d'importance les variables corrélées avec la croissance du produit. On s'aperçoit alors que le modèle statistique implicite dans ce genre d'exercice n'est pas des plus clair. Comme le note Durlauf [44], les méthodes de sélection de "modèle" posent encore problème même d'un point de vue purement statistique.

reposant sur la diminution des rendements du capital ne tient plus, car dans ce cadre le capital se répartit instantanément entre les économies de façon à égaliser les prix des facteurs (cette égalisation peut être conditionnelle). Malgré cela, Glachant et Vellutini [64] montrent[25] que ce modèle produit une dynamique de transition que l'on ne peut véritablement associer à de la convergence conditionnelle car la distribution de long terme des richesses s'explique par sa distribution initiale.

Du fait de l'égalisation des rendements la progression de la richesse totale est identique dans toutes les économies. Mais les deux auteurs établissent que la distribution de la richesse financière (*ie* du capital) évolue si l'économie mondiale est caractérisée par une dynamique de transition. En particulier, si cette dynamique résulte en un accroissement de la richesse totale inférieur à l'accroissement de la richesse financière, alors la distribution des richesses financières se contracte[26] tant que l'économie agrégée n'a pas atteint son état stationnaire. Cette transition de la distribution de la richesse financière s'explique par une modification de la composition de la richesse totale lors de la transition de l'économie agrégée. Dans la mesure de l'élasticité de substitution entre capital et travail, lorsque l'économie agrégée accumule du capital la demande pour le travail se contracte entraînant ainsi une diminution des flux anticipés de salaire, *ie* une baisse de la part non financière de la richesse totale. Afin d'assurer l'égalité des taux de croissance de la richesse totale les économies initialement faiblement dotées en capital doivent plus accumuler que les économies initialement mieux dotées. Cet effet de recomposition de la richesse totale explique ainsi le rapprochement des économies en termes de richesse financière, mais aussi en termes de produit national brut. Le processus de convergence est bloqué lorsque la croissance de la richesse financière égalise la croissance de la richesse totale, c'est à dire lorsque l'économie évolue le long de sa trajectoire de long terme. Ainsi une part des inégalités initiales demeure dans la distribution limite.

On voit alors que la dynamique de convergence n'est pas stable dans la chronologie et qu'elle est intimement liée à la croissance de l'économie agrégée. Adjemian, Glachant et Vellutini [6] utilisent ce lien dans un voisinage de l'état stationnaire et montrent que les vitesses de convergence à la Barro es-

[25]Voir aussi Ventura [151].

[26]Ils dérivent un résultat analogue pour le produit.

timées sur l'échantillon des états américains (voir le graphique [1.2, en page 43]) suivent de près l'évolution de l'économie agrégée. Une croissance rapide de l'économie agrégée s'accompagne d'une réduction des inégalités. Il s'agit donc d'une explication alternative à l'instabilité chronologique du processus de convergence. La décélération du processus de convergence s'explique alors par le rapprochement de l'économie agrégée de sa cible de long terme.

L'adoption d'une perspective radicalement différente sur le degré d'ouverture des économies rend ce modèle presque aussi insatisfaisant que celui considéré dans cette section. L'hypothèse d'une mobilité non contrainte du capital, et donc de marché parfait, est aussi forte que l'hypothèse d'un monde composé d'économies autarciques. Il nous manque un modèle intermédiaire avec des économies imparfaitement intégrées.

3 Estimation des vitesses de Convergence

Dans la section [2] nous avons vu qu'une régression à la Barro ne peut donner beaucoup d'information sur la vitesse de convergence néo-classique. Elle peut tout au plus donner une idée sur les sources de la convergence et une information sur le degré de persistance des chocs de productivité.

Dans cette section on propose une procédure d'estimation des deux vitesses de convergence, en prenant le modèle de croissance néo-classique *au pied de la lettre*. Pour cela on considère un modèle de croissance optimale stochastique proche du modèle présenté dans la section [2] et on exploite pleinement la dimension temporelle. Une idée récurrente parmi les défenseurs acharnés de l'*approche classique* est qu'il n'est pas profitable d'exploiter pleinement la dimension temporelle car les variations de court terme sont brouillées par les cycles d'affaire. Par exemple, Temple [149, page 27] écrit,

> "*My own feeling is that time series econometricians are usually mistaken in believing that there is much to be gained by moving to the use of annual data, even in panels. Inevitably the short run variation in growth rates is dominated by business cycle effects, not by changes in fundamental long run growth prospects. Only the long time averages of growth rates, compared in the cross-section or using a panel, can allow us to address the determinants of long run growth with any degree of confidence.*"

L'avis donné par Temple peut se retrouver chez de nombreux auteurs affiliés à l'*approche classique*. Par exemple, Shioji [131] développe le même genre d'argument pour affirmer que la régression en coupe est préférable au panel. Si on peut estimer une vitesse de convergence de façon convergente lorsque l'horizon T tend vers l'infini avec une régression en coupe, rien n'assure que cette vitesse de convergence soit liée au mécanisme que l'on cherche à identifier (les rendements décroissants dans le capital). Le contenu de la section [2] montre plutôt le contraire. Dans cette section on propose une approche empirique cohérente avec le processus générateur des données pour estimer les vitesses de convergence. Contrairement à l'*approche classique*, on n'exploite pas la dimension individuelle mais la dynamique de court terme pour évaluer les vitesses de convergence.

A chaque période l'économie est perturbée par un choc technologique qui se diffuse imparfaitement. Le paramètre qui contrôle le rythme de cette diffusion est la vitesse de convergence technologique. On montre alors, en suivant Campbell [27], que le produit par travailleur est un processus ARIMA. On déduit les deux vitesses de convergence à partir des racines réelles de la partie autorégressive. Ainsi il est théoriquement possible d'estimer les vitesses de convergence en exploitant pleinement la dimension temporelle. Plus généralement, si le modèle considéré contient p variables d'état, alors on devrait pouvoir estimer p vitesses de convergence (si le modèle exhibe un équilibre déterminé alors il y a autant de bras stables, *ie* de vitesses de convergence, que de variables d'état) en considérant les racines de la partie autorégressive du processus déterminant la loi d'évolution (induite par le modèle) du produit par tête.

L'avantage de cette approche est que l'on ne suppose pas l'existence d'une tendance de progrès technique commune dans la coupe, ni l'homogénéité des paramètres fondamentaux. Ces hypothèses, implicitement admises par l'*approche classique*, sont empiriquement difficiles à justifier. Notre approche peut paraître assez contre-intuitive, au regard de la littérature, car on obtient les vitesses de convergence en considérant une seule économie. Il faut rappeler que dans le modèle de croissance néo-classique, les économies ne convergent pas les unes vers les autres mais plutôt vers une cible que l'on suppose commune (ou vers des cibles parallèles). Il n'est donc pas nécessaire d'exploiter l'information dans la dimension individuelle pour estimer les vitesses de convergence.

La section [3.1] établit qu'il devrait être possible d'estimer de façon convergente les vitesses de convergence technologique et néo-classique en exploitant pleinement les séries temporelles. La section [3.2] propose une application sur données américaines.

3.1 Le modèle de croissance stochastique

Dans cette section on écrit un modèle de croissance stochastique et on le résout *à la main* dans un voisinage de l'état stationnaire. Ceci permet d'obtenir une forme réduite décrivant l'évolution du produit par travailleur.

La technologie de production est de type néo-classique, le produit Y_t est obtenu à partir du capital K_t et du travail efficace A_tL_t, $Y_t = F(K_t, A_tL_t)$. On suppose que le terme d'efficacité Harrod-neutre est de la forme[27],

$$\begin{aligned} A_{t+1} &= A_t(1+g)e^{x_{t+1}} \\ x_{t+1} &= \rho x_t + \varepsilon_{t+1} \end{aligned}$$

avec $g > 0$ le taux de croissance espéré de la technologie Harrod neutre, et $\varepsilon_t \sim\ iid(0, \sigma_\varepsilon^2)$. On notera $X_t = e^{x_t}$. Le A est tendanciel au sens déterministe (son espérance est une fonction croissante du temps) et au sens stochastique (sa variance est une fonction croissante du temps). Le paramètre ρ mesure le rythme de la diffusion d'un choc technologique. Une façon alternative de voir le A est d'utiliser la décomposition tendance – cycle de Beveridge-Nelson[28]. Si on note $a_t = \log A_t$ alors on peut écrire $a_t = a_t^{\mathcal{T}} + a_t^{\mathcal{C}}$, avec

$$\begin{aligned} a_t^{\mathcal{T}} &= a_{t-1}^{\mathcal{T}} + g + \frac{\varepsilon_t}{1-\rho} \\ a_t^{\mathcal{C}} &= \rho a_{t-1}^{\mathcal{C}} - \frac{\rho}{1-\rho}\varepsilon_t \end{aligned}$$

L'économie n'est pas toujours sur la tendance $a_t^{\mathcal{T}}$, que l'on peut interpréter comme la tendance de progrès technique. Selon le signe du terme cyclique $a_t^{\mathcal{C}}$, elle gravite autour de cette tendance. Si $a_t^{\mathcal{C}}$ est négatif, alors cette économie est temporairement inefficace, dans le sens où elle ne parvient pas à mettre en œuvre la technologie. Dans ce cadre, contrairement au modèle envisagé dans la section précédente, l'économie peut temporairement se placer au-delà de la meilleure pratique, lorsque $a_t^{\mathcal{C}}$ est positif.
On suppose que la loi d'évolution du nombre de travailleur est déterministe. Ainsi $L_t = (1+n)^t$, avec $n > 0$, représente le nombre de travailleurs à la date t. De la même façon que dans la section [2] la fonction de production est caractérisée par deux paramètres : la part du capital dans la valeur ajoutée (α) et l'élasticité de substitution entre capital et travail (ε).

[27]Nous pourrions envisager une spécification différente du processus $\{x_t, t \in \mathbb{Z}\}$.
[28]Voir Hamilton [70, pages 504-505].

L'utilité intertemporelle espérée conditionnellement à l'information disponible à la date t est donnée par

$$\mathbb{E}_t\left[\sum_{\tau=0}^{\infty}\beta^{*\tau}u(C_{t+\tau})\right]$$

où C_t est le niveau de la consommation à la date t, $u(\bullet)$ est la fonction d'utilité instantanée CRRA (on note σ l'inverse de l'élasticité de substitution intertemporelle). Le planificateur social maximise l'utilité espérée,

$$\begin{aligned}\underset{\{C_{t+\tau},\tau\in\mathbb{N}\}}{\mathcal{M}ax}\ \mathbb{E}_t & \left[\sum_{\tau=0}^{\infty}\beta^{*\tau}u(C_{t+\tau})\right] \\ & \underline{sc}\ K_{t+1}=F(K_t,A_tL_t)-C_t+(1-\delta)K_t \\ & A_t=A_{t-1}(1+g)X_t \\ & \log X_t=\rho\log X_{t-1}+\varepsilon_t \\ & K_0,\ A_0 \text{ donnés.}\end{aligned} \tag{1.17}$$

où K_t et δ sont respectivement le stock de capital physique et le taux de dépréciation constant du capital.

En écrivant la fonction valeur associée à ce programme dynamique (voir la section [5, en page 113]) on obtient l'équation d'Euler,

$$u'(C_t)=\beta^*\mathbb{E}_t\left[u'(C_t)R_{t+1}\right] \tag{1.18}$$

où $R_{t+1}\equiv 1+r_t=F_1(K_{t+1},A_{t+1}L_{t+1})+1-\delta$ est le facteur d'intérêt réel à la date $t+1$. L'équation d'Euler décrit la trajectoire de consommation optimale et caractérise, avec l'équation d'accumulation du capital, la dynamique de l'économie. La condition de transversalité pose une contrainte terminale sur cette dynamique, elle est donnée par

$$\lim_{s\to\infty}\beta^{*s}\mathbb{E}_t\left[\lambda_s K_{s+1}\right]=0 \tag{1.19}$$

où λ_s est la valeur marginale anticipée d'une unité supplémentaire de capital physique à la date $s+1$. Cette condition stipule que la détention de capital n'est pas valorisée asymptotiquement. En utilisant l'homogénéité de degré un de la fonction de production et en notant que l'utilité marginale est homogène de degré $-\sigma$ (puisque la fonction d'utilité instantanée est de type CRRA), on peut réécrire la dynamique non linéaire en termes de variables par tête efficace[29],

$$\begin{aligned}(1+n)(1+g)X_{t+1}\hat{k}_{t+1} &= F(\hat{k}_t,1) - \hat{c}_t + (1-\delta)\hat{k}_t & (1.20)\\ 1 &= \beta\mathbb{E}_t\left[\frac{u'(\hat{c}_{t+1})}{u'(\hat{c}_t)}X_{t+1}^{-\sigma}\left\{F_1(\hat{k}_{t+1},1)+1-\delta\right\}\right] & (1.21)\end{aligned}$$

avec $\beta \equiv \beta^*\left[(1+n)(1+g)\right]^{-\sigma}$. La dynamique linéarisée autour de l'état stationnaire (voir la section [5]) est donnée par[30],

$$\tilde{k}_{t+1} = (\omega_1+\omega_2)\tilde{k}_t + \left(1-\omega_1-\frac{\omega_2}{\alpha}\right)\tilde{c}_t - x_{t+1} \tag{1.22}$$

et

$$\mathbb{E}_t\left[\Delta\tilde{c}_{t+1}\right] = -\omega_3\mathbb{E}_t\left[\tilde{k}_{t+1}\right] - \mathbb{E}_t\left[x_{t+1}\right] \tag{1.23}$$

avec,

$$\begin{aligned}\omega_1 &= \frac{1-\delta}{(1+n)(1+g)}\\ \omega_2 &= \frac{r^*+\delta}{(1+n)(1+g)}\\ \omega_3 &= \sigma^{-1}\frac{r^*+\delta}{1+r^*}\frac{1-\alpha}{\varepsilon}\end{aligned}$$

Notons que l'on a nécessairement $\omega_1+\omega_2>1$ si la condition de transversalité (1.19) est vérifiée. Pour résoudre ce modèle à anticipation rationnelle, c'est à dire pour écrire chaque variable en fonction du passé, on procède, comme Campbell [27], par la méthode des coefficients indéterminés. On écrit la variable de décision ($\tilde{c}_t$) en fonction des deux variables d'état du modèle linéarisé autour de l'état stationnaire ($\tilde{k}_t$ et x_t),

$$\tilde{c}_t = \eta_{ck}\tilde{k}_t + \eta_{cx}x_t \tag{1.24}$$

où η_{ck} et η_{cx} représentent respectivement les élasticités partielles[31] de la

[29] On pose $\hat{k}_t = K_t/(A_tL_t)$.
[30] On pose $\tilde{k}_t = (\hat{k}_t - \hat{k}^*)/\hat{k}^*$.
[31] On suit la terminologie de Campbell [27].

consommation relativement au capital et à la productivité. En substituant (1.24) dans l'équation linéarisée d'évolution du capital (1.22), il vient

$$\tilde{k}_{t+1} = \eta_{kk}\tilde{k}_t + \eta_{kx}x_t - x_{t+1} \tag{1.25}$$

avec

$$\begin{aligned} \eta_{kk} &= \omega_1 + \omega_2 + \left(1 - \omega_1 - \frac{\omega_2}{\alpha}\right)\eta_{ck} \\ \eta_{kx} &= \left(1 - \omega_1 - \frac{\omega_2}{\alpha}\right)\eta_{cx} \end{aligned}$$

Les élasticités partielles du capital en $t+1$ par rapport au capital en t et à la productivité en t. Ces deux paramètres sont fondamentaux pour décrire la dynamique du modèle. En effet, $(1 - \eta_{kk})$ correspond à la vitesse de convergence néo-classique. Le paramètre η_{kx} est essentiel dans l'évaluation de la corrélation entre capital et productivité. En substituant (1.25) et la *policy rule* (1.24) dans l'équation d'Euler linéarisée (1.23), on obtient par identification deux équations qui déterminent les élasticités partielles η_{ck} et η_{cx},

$$Q_2\eta_{ck}^2 + Q_1\eta_{ck} + Q_0 = 0 \tag{1.26}$$

et

$$\eta_{cx} = \rho\frac{\omega_3 + \eta_{ck} - 1}{(\omega_3 + \eta_{ck})\,Q_2 + \rho - 1} \tag{1.27}$$

avec

$$\begin{aligned} Q_2 &= 1 - \omega_1 - \frac{\omega_2}{\alpha} \\ Q_1 &= \omega_1 + \omega_2 - 1 + \omega_3\left(1 - \omega_1 - \frac{\omega_2}{\alpha}\right) \\ Q_0 &= \omega_3(\omega_1 + \omega_2) \end{aligned}$$

En notant que pour toute paramétrisation du modèle (vérifiant la condition de transversalité) on a $Q_2 < 0$ et $Q_0 > 0$, on voit immédiatement que l'équation (1.26) a deux racines réelles de signes opposés. On ne peut assigner à η_{ck} la racine négative car cela engendrerait une dynamique explosive (on aurait alors $\eta_{kk} > 1$). On a donc,

$$\eta_{ck} \quad = \quad -\frac{1}{2Q_2}\left\{Q_1 + \sqrt{Q_1^2 - 4Q_0Q_2}\right\}$$

Notons que η_{ck} et la vitesse de convergence néo-classique sont indépendants de ρ. Comme dans la section précédente, dans un voisinage de l'état stationnaire, on a orthogonalité entre la vitesse de convergence technologique $(1-\rho)$ et la vitesse de convergence néo-classique $(1 - \eta_{kk})$.

On peut maintenant décrire la loi d'évolution du taux de croissance du produit par travailleur. Pour nous résumer, on a sait que dans un voisinage de l'état stationnaire, on a

$$\tilde{y}_t = \alpha\tilde{k}_t \tag{1.28}$$

$$x_t = \rho x_{t-1} + \varepsilon_t \tag{1.29}$$

$$\tilde{k}_{t+1} = \eta_{kk}\tilde{k}_t + \eta_{kx}x_t - x_{t+1} \tag{1.30}$$

En notant que le taux de croissance du produit par travailleur efficace en déviation à l'état stationnaire est donné par,

$$\Delta\tilde{y}_{t+1} = \Delta\log y_{t+1} - g - x_{t+1}$$

où $y_t = Y_t/L_t$ et $\Delta = 1 - L$ représente l'opérateur différence première, on obtient une première expression du taux de croissance du produit par travailleur,

$$\Delta\log y_{t+1} = \alpha\Delta\tilde{k}_{t+1} + g + \frac{\varepsilon_{t+1}}{1-\rho L} \tag{1.31}$$

En multipliant (1.31) par $(1 - \eta_{kk}L)(1 - \rho L)$, on obtient,

$$(1 - \eta_{kk}L)(1 - \rho L)\Delta\log y_{t+1} = \alpha\Delta\tilde{k}_{t+1} + \breve{g} + (1 - \eta_{kk}L)\varepsilon_{t+1} \tag{1.32}$$

avec $\breve{g} = (1 - \eta_{kk})(1 - \rho)g$. En multipliant l'équation (1.30) en différence première par $(1 - \rho L)$, il vient,

$$(1 - \eta_{kk}L)(1 - \rho L)\Delta\tilde{k}_{t+1} = \eta_{kx}\Delta\varepsilon_t - \Delta\varepsilon_{t+1}$$

En substituant cette dernière équation dans l'équation (1.32), on obtient facilement la loi d'évolution du produit par travailleur cohérente avec le modèle,

$$(1-\eta_{kk}L)(1-\rho L)\Delta\log y_{t+1} = \breve{g} + u_{t+1} + \frac{\alpha\eta_{kx}+\alpha-\eta_{kk}}{1-\alpha}u_t - \frac{\alpha\eta_{kx}}{1-\alpha}u_{t-1}$$

ou de façon équivalente,

$$\begin{aligned}\Delta\log y_{t+1} &= \breve{g} + (\eta_{kk}+\rho)\Delta\log y_t - \eta_{kk}\rho\Delta\log y_{t-1} + u_{t+1} \\ &+ \frac{\alpha\eta_{kx}+\alpha-\eta_{kk}}{1-\alpha}u_t - \frac{\alpha\eta_{kx}}{1-\alpha}u_{t-1} \qquad (1.33)\end{aligned}$$

avec $u_t = (1-\alpha)^{-1}\varepsilon_t$. Ainsi le taux de croissance du produit par travailleur est un processus ARMA(2,2). Les racines du polynôme caractéristique associé à la partie autorégressive du processus doivent être réelles et correspondent aux paramètres ρ et η_{kk}. Ce résultat donne une piste intéressante pour estimer les vitesses de convergence néo-classique $(1-\eta_{kk})$ et de rattrape technologique $(1-\rho)$. Si on observe le produit par travailleur , on peut identifier les paramètres η_{kk} et ρ. En considérant la partie MA, on devrait aussi pouvoir identifier l'élasticité partielle η_{kx} qui indique l'effet du choc de productivité sur le capital par travailleur efficace, la part du capital dans la valeur ajoutée et la variance de l'innovation ε_t.

Il convient de noter que l'on a raisonné sur la base d'un comportement d'offre de travail exogène. L'identification des vitesses de convergence néo-classique et technologique à partir des racines de la partie autorégressive est toujours possible dans un cadre plus général. Si on amende le modèle avec un comportement d'offre de travail endogène (en adoptant une fonction d'utilité séparable ou non entre loisir et consommation), la dynamique du taux de croissance du produit par travailleur est toujours un processus ARMA(2,2)[32] et l'interprétation des racines de la partie autorégressive ne change pas. Évidemment l'expression de la vitesse de convergence $(1-\eta_{kk})$ est plus complexe.

Finalement il faut noter que la modélisation de $\{x_t, t\in\mathbb{Z}\}$ est totalement *ad-hoc*. Le choix d'un processus autorégressif d'ordre un est simplement motivé par la simplicité et l'usage. On pourrait envisager un processus AR(p),

[32] Voir Campbell [27].

et montrer[33] que le taux de croissance du produit par travailleur est alors un processus ARMA$(p+1, p+1)$. On peut aussi montrer que l'expression de la vitesse de convergence néo-classique ne dépend pas de la forme du processus exogène de productivité.

3.2 Estimation des vitesses de convergence

Dans cette section on propose une estimation des deux vitesses de convergence[34]. Nous devons estimer un processus de la forme,

$$\Delta y_{i,t} = \tilde{g}_i + \phi_{1,i}\Delta y_{i,t-1} + \phi_{2,i}\Delta y_{i,t-2} + w_{i,t} + \theta_{1,i}w_{i,t-1}\theta_{2,i}w_{i,t-2} \qquad (1.34)$$

Pour la calibration retenue (voir l'annexe), on a $\phi_{1,i} = 1,9073$, $\phi_{2,i} = -0,9088$, $\theta_{1,i} = -1,2847$ et $\theta_{1,i} = 0,3869$. La vitesse de convergence néoclassique est alors 7,3%. Notons que cette approche, à l'instar de l'approche par les séries temporelles de la convergence, impose que les données ne soient pas caractérisées par une dynamique de transition importante (afin que la linéarisation du modèle ne pose pas de problème). Ainsi on ne pourrait appliquer cette approche pour estimer les vitesses de convergence des économies européennes sur des données d'après seconde guerre mondiale.

On considère le taux de croissance du produit par travailleur aux États Unis entre 1980 :01 et 2002 :02[35], les figures [1.14, en page 117] et [1.15, en page 118] représentent ces données en niveau et en croissance. La mise en œuvre d'un test de Schmidt–Phillips sur le logarithme du niveau du produit réel par travailleur montre qu'il s'agit de la réalisation d'un processus non stationnaire à racine unitaire. La statistique de student estimée est -1,4166 ; on accepterait l'hypothèse nulle de racine unitaire autour d'une tendance linéaire même avec un seuil de 80%. Ainsi l'acceptation de cette hypothèse est

[33] Voir Campbell [27]. Il faut alors redéfinir la *policy rule* en ajoutant des retards sur la variable d'état x_t.

[34] On ne teste pas le modèle et on ne tente pas une identification des paramètres structurelles qui apparaissent dans la partie moyenne mobile.

[35] Les données sont fournies par le BLS. Ces données sont trimestrielles et corrigées des variations saisonnières. Les taux de croissance sur un trimestre sont annualisés. Ces données excluent l'activité agricole et ne considèrent que la production privée.

forte. Il apparaît donc légitime de considérer le modèle en différence première, en raisonnant sur le taux de croissance du produit par travailleur. En utilisant les méthodes du coin (voir Gourieroux et Monfort [65, pages 191-193]) et du triangle (*Extended Sample Autocorrelation Function*, voir Tsay et Tiao [150]) on peut sélectionner les modèles ARMA suivants pour le taux de croissance du produit par travailleur,

méthode du coin	méthode du triangle
ARMA(2,2)	ARMA(3,3)
ARMA(5,0)	ARMA(0,5)
ARMA(0,5)	ARMA(1,5)
	ARMA(2,5)

Le processus générateur des données théorique nous conduit à ne considérer que les spécifications ARMA(2,2) et ARMA(3,3). La dernière spécification correspond au cas où le processus $\{x_t, t \in \mathbb{Z}\}$ est un AR(2). Le fait que ces méthodes automatiques de sélection de modèle envisagent des spécifications théoriquement justifiables ne prouve aucunement la validité du modèle théorique, mais est rassurant. L'estimation du modèle ARMA(2,2) par le maximum de vraisemblance aboutit à l'estimation de racines complexes pour la partie autorégressive. Ce modèle est inexploitable, car le modèle prédit que les racines de la partie AR doivent être réelles. L'estimation du modèle ARMA(3,3) par le maximum de vraisemblance[36] donne,

$$\begin{aligned}\Delta y_t = \; & \underset{[18{,}53]}{2,26}\Delta y_{t-1} - \underset{[-10{,}52]}{2,11}\,\Delta y_{t-2} + \underset{[7{,}04]}{0,79}\Delta y_{t-3} \\ & + \widehat{\varepsilon}_t - \underset{[-21{,}57]}{2,25}\,\widehat{\varepsilon}_{t-1} + \underset{[12{,}90]}{2,16}\widehat{\varepsilon}_{t-2} - \underset{[-9{,}75]}{0,91}\,\widehat{\varepsilon}_{t-3}\end{aligned} \tag{1.35}$$

La partie autorégressive est caractérisée par deux racines complexes conjuguées $(0,69 \pm 0,65i)$ et une racine réelle $(0,87)$. Les racines complexes sont liées au processus $\{x_t, t \in \mathbb{Z}\}$[37]. La norme des racines complexes permet

[36] Les conditions initiales sont déterminées par la calibration du modèle. On a centré les données pour estimer le modèle, le taux de croissance moyen du produit par travailleur entre 1980 :01 et 2002 :02 est 1,71%.

[37] Le modèle présenté plus haut ne peut générer une dynamique à racines complexes. Ainsi les racines complexes ne peuvent provenir que du processus de productivité exogène (une AR(2)).

donc d'obtenir la vitesse de convergence technologique, la vitesse de convergence néo-classique est obtenue à partir de la racine réelle. Ainsi la vitesse de convergence néo-classique estimée est 12,8% et la vitesse de convergence technologique, définie par la norme des racines complexes, est 5%. On obtient une vitesse de convergence technologique inférieure à la vitesse de convergence néo-classique.

Malheureusement, la partie moyenne mobile contient une racine unitaire réelle et deux racines complexes. Si l'estimation de la partie autorégressive semble plutôt satisfaisante, on est confronté à un problème pour la partie moyenne mobile. Cela peut s'expliquer par le fait que, pour la calibration retenue, le modèle prédit une racine proche de l'unité dans la partie MA. Si, de plus, la variance de l'innovation de la tendance stochastique est petite, *ie* si on est proche d'un cas de sur-différenciation, alors l'estimation d'une racine unitaire dans la partie MA est moins étonnante. En effet, Sargan et Barghava [129][38] argumentent que l'estimation d'une racine unitaire dans la partie MA d'un processus ne veut pas nécessairement dire que le processus est sur-différencié. La vraisemblance pourrait bien avoir un maximum local (voire global) avec une racine unitaire dans la partie MA, même si le vrai processus n'est pas caractérisé par une racine unitaire dans la partie MA. Ils reportent des simulations où il apparaît que même si la vraie racine est inférieure à l'unité, la probabilité pour que la vraisemblance soit maximisée avec une racine unitaire dans la partie MA est forte (d'autant plus que le modèle contient des variables tendancielles ou fortement auto-corrélées. Notons que pour la calibration retenue nous ne sommes pas très éloigné d'une racine unitaire dans la partie MA.
On pourrait aider l'algorithme d'optimisation en spécifiant les contraintes induites par le modèle sur la partie MA.

3.3 Conclusion

Dans cette section nous avons proposé une stratégie d'estimation des vitesses de convergence néo-classique et technologique cohérente avec un modèle de croissance et exploitant pleinement la dimension temporelle. Cette

[38] voir aussi Maddala et kim [94, page 116] ou, mieux, Tanaka [148, chapitre 8].

démarche a le mérite de ne pas imposer des restrictions d'homogénéité dans la coupe des économie (puisque nous n'exploitons pas cette dimension), contrairement à l'*approche classique*. Il ne s'agit pas de la première tentative d'estimation de vitesse de convergence exploitant pleinement la dimension temporelle.

Evans [50] propose une stratégie d'estimation qui n'impose pas l'homogénéité des vitesses de convergence. Il considère un échantillon de N économies qui convergent, au sens où les écarts de produit par tête sont stationnaires, et suppose que le produit par tête d'une économie en déviation à la moyenne individuelle est un processus de la forme,

$$\lambda_n(L)(y_{n,t} - \bar{y}_{.,t}) = \delta_n + \varepsilon_{n,t} \tag{1.36}$$

où $\lambda_n(L) = 1 - \sum_{i=1}^{p+1} \lambda_{i,n} L^i$ est un polynôme retard vérifiant $\lambda_n(z) = 0$ implique $z > 1$, δ_n est une constante liée au niveau de long terme de l'économie n par rapport aux autres économies, $\varepsilon_{n,t}$ est un bruit blanc de variance σ_n^2 indépendant dans la coupe, $t = 1, ..., T$. La vitesse de convergence de l'économie n est sa vitesse de retour au niveau de long terme $\delta_n / \lambda_n(1)$. Cette vitesse, notée r_n, correspond à la plus grande racine du polynôme caractéristique $z^{p+1}\lambda_n(z^{-1})$, on note $r_n = 1 - z_n^*$. Evans, en suivant Stock [145], propose d'utiliser la théorie des processus presque intégrés[39] (*local to unity asymptotic*) pour obtenir une estimation non biaisée de la vitesse de convergence, *ie* de la racine z_n^*, et un intervalle de confiance. Il ne s'agit pas de dire que la vitesse de convergence tend vers zéro lorsque la taille de l'échantillon tend vers l'infini, mais plutôt d'utiliser une théorie asymptotique susceptible de fournir une meilleure approximation à distance finie. Il réécrit le processus

[39] Pour une introduction sur le sujet des processus presque intégrés on peut se reporter à Phillips [109]. Brièvement un processus presque intégré est de la forme $y_t = \rho y_{t-1} + \varepsilon_t$ $(t = 1, ..., T)$ avec $\rho = e^{\frac{c}{T}} = 1 + \frac{c}{T} + o(T^2)$. Lorsque la taille de l'échantillon (T) tend vers l'infini, le paramètre autorégressif tend vers l'unité. Si le paramètre de localisation est négatif le processus est très persistant et tend vers une marche aléatoire pure. Cette théorie asymptotique peut servir à étudier la puissance des tests de racine unitaire contre des alternatives locales, à construire des intervalles de confiance pour des processus fortement corrélés...

générateur des données (1.36) de la façon suivante,

$$\Delta(y_{n,t} - y_{.,t}) = \delta_n + \rho_n(y_{n,t-1} - y_{.,t-1}) + \sum_{i=1}^{p} \varphi_{n,i}\Delta(y_{n,t-i} - y_{.,t-i}) + \varepsilon_{n,t} \quad (1.37)$$

avec $\rho_n = \sum_{j=1}^{p+1} -1 < 0$ et $\varphi_{n,i} = -\sum_{j=i+1}^{p+1} \lambda_{n,j}$. Evans suppose que la vitesse de convergence est telle que Tr_n est constant lorsque T tend vers l'infini, c'est à dire que la racine z_n^* tend vers l'unité à la même vitesse que T diverge. Evans fait appel à une asymptotique locale car celle-ci est susceptible de fournir une meilleure approximation à distance finie de la distribution de l'estimateur de la plus grande racine du polynôme caractéristique (si cette racine est proche de l'unité). Stock [145] montre que le student associé à l'estimateur (des MCO) de ρ_n tend vers une distribution limite (une fonctionnelle de processus d'Ornstein–Uhlenbeck) où Tr_n[40] apparaît comme l'unique paramètre de nuisance. On note $\hat{\tau}_n \Rightarrow \mathcal{L}(Tr_n)$. Les quantiles de cette distribution sont généralement une fonction continue et bijective de ce paramètre de nuisance. Ainsi, il est possible de déduire un estimateur de Tr_n en se concentrant par exemple sur la médiane de la distribution,

$$\tilde{r}_n = \frac{1}{T}\mathcal{L}^{-1}_{q=0.50}(\hat{\tau}_n) \quad (1.38)$$

Cet estimateur de la vitesse de convergence est non biaisé dans le sens où il y a autant de chance pour qu'il soit supérieur à r_n ou inférieur à r_n. Ensuite on peut simplement construire un intervalle de confiance pour cet estimateur en considérant, par exemple, $\frac{1}{T}\mathcal{L}^{-1}_{q=0.05}(\hat{\tau}_n)$ et $\frac{1}{T}\mathcal{L}^{-1}_{q=0.95}(\hat{\tau}_n)$. Evans [50] applique cette procédure d'estimation sur les états américains. Les résultats sont assez décevants car les intervalles de confiance sont très larges. Pour treize états, il estime une vitesse de convergence négative. Pour trente trois états la vitesse de convergence n'est pas significativement différente de zéro. Ces résultats ne sont pas réellement étonnants car il existe peu de résultats empiriques favorables à la convergence des états américains[41]. Néanmoins, Evans note que la vitesse de convergence moyenne[42] est de 15,5%, une estimation large-

[40]Il s'agit du paramètre de localisation définissant le processus presque intégré.

[41]Evans et Karras [52] obtiennent un résultat de convergence des états américains, mais on verra dans le chapitre [2] que ce résultat est sujet à caution.

[42]On peut s'interroger sur le sens qu'il peut y avoir à considérer une moyenne. Evans développe une approche d'estimation en affirmant qu'il est important de ne pas supposer

ment supérieure à celle reportée par les tenants de l'*approche classique*. Puis il confronte cette estimation au modèle de croissance néo-classique et conclut qu'il n'est pas nécessaire d'envisager une conception élargie du capital physique pour que le modèle soit cohérent avec les données.

L'approche d'Evans [50] est radicalement différente de la notre et souffre de plusieurs défauts. Premièrement, on peut s'interroger sur la pertinence du processus étudié. Evans considère le produit par tête en déviation au produit par tête moyen. Si les N économies convergent alors ces N processus doivent être stationnaires. Si une seule économie diverge alors les N processus considérés par Evans doivent être non stationnaires[43]. Lorsque l'on estime une vitesse de convergence négative ou non significativement différente de zéro (divergence) pour certains états et une vitesse de convergence positive pour d'autres, il y a une sorte de contradiction. Il serait plus pertinent de considérer des écarts à un état pour autoriser la divergence de certains états et la convergence d'autres états. Evans adopte une approche intermédiaire en exploitant partiellement l'information dans la coupe (via la moyenne individuelle du produit par tête), ce qui ne contribue à faciliter l'interprétation de ses résultats. Deuxièmement, on peut s'interroger sur la pertinence de la comparaison avec le modèle de croissance néo-classique. Cette comparaison suppose implicitement que la vitesse de convergence estimée par cette procédure d'estimation identifie la vitesse de convergence néo-classique. Rien ne permet d'étayer cette hypothèse. D'autant plus que la spécification AR adoptée par Evans est incompatible avec un modèle de croissance optimale stochastique. Dans ce sens, le processus considéré par Evans est mal spécifié, il ne peut dire quelle vitesse de convergence il identifie.

l'homogénéité des vitesses de convergence, mais finalement il en vient à considérer une moyenne des vitesses de convergence. Autant utiliser directement une stratégie d'estimation en panel de la vitesse de convergence. On a adopté cette approche en considérant la procédure d'estimation du paramètre de localisation proposé par Phillips et Moon [113]. Cette stratégie d'estimation en panel est plus simple que celle proposée par Stock [145], car en profitant de la dimension panel on obtient directement un estimateur convergent du paramètre de localisation (s'il est homogène dans la dimension individuelle). Une estimation sur les états américains donne une vitesse de convergence négative. Encore un fois, un tel résultat n'est pas étonnant dans la mesure où ces économies ne convergent pas au sens des séries temporelles, *ie* les écarts de revenus ne sont pas stationnaires en niveau.

[43] Voir Evans et Karras [52] et le chapitre [2].

Notre approche a le double avantage de ne pas imposer l'homogénéité des vitesses de convergence et d'être cohérente avec un modèle économique. Cependant, il convient de noter que notre interprétation des racines de la partie autorégressive comme les vitesses de convergence néo-classique et technologique n'est valable que si les données sont effectivement cohérentes avec le modèle. Nous ne pourrons nous faire une idée sur ce dernier point avant d'obtenir des estimations plus satisfaisantes pour la partie MA. On pourrait, par exemple, envisager d'utiliser une approche bayésienne pour estimer le processus ARMA.

4 Annexe de la section 2

4.1 Dérivation des résultats de la section 2.1

En écrivant le Hamiltonien associé au programme du planificateur social et les conditions de Pontryagin, on obtient la dynamique des variables par travailleur sous la forme d'une système de trois équations,

$$\begin{cases} \dot{c}(t) = c(t)\sigma(t)^{-1}\left[A(t)F_1(k(t),T(t)) - \delta - \rho\right] \\ \dot{k}(t) = A(t)F(k(t),T(t)) - (\delta+n)k(t) - c(t) \\ \dot{A}(t) = \phi(1-A(t)) \end{cases} \tag{1.39}$$

En posant $\hat{k}(t) = k(t)/T(t)$ et $\hat{c}(t) = c(t)/T(t)$ les variables "détrendées" du progrès technologique, on peut réécrire la dynamique (1.39) de la façon suivante,

$$\begin{cases} \dot{\hat{c}}(t) = \hat{c}(t)\left\{\sigma(t)^{-1}\left[A(t)F_1(\hat{k}(t),1) - \delta - \rho\right] - x\right\} \\ \dot{\hat{k}}(t) = A(t)F(\hat{k}(t),1) - (\delta+n+x)\hat{k}(t) - \hat{c}(t) \\ \dot{A}(t) = \phi(1-A(t)) \end{cases}$$

L'état stationnaire de cette dynamique est défini par,

$$\begin{cases} \hat{c}^* = F\left(\hat{k}^*,1\right) - (\delta+n+x)\hat{k}^* \\ F_1\left(\hat{k}^*,1\right) = \sigma x + \rho + \delta \\ A^* = 1 \end{cases}$$

On vérifie facilement que la contrainte de transversalité est vérifiée dès lors que l'inégalité suivante est vérifiée,

$$\rho - n > (1-\sigma)x$$

En posant $\tilde{x}(t) = \log \hat{x}(t) - \log \hat{x}^*$, la dynamique linéarisée autour de l'état stationnaire s'écrit sous forme matricielle de la façon suivante,

$$\begin{pmatrix} \dot{\tilde{k}}(t) \\ \dot{\tilde{c}}(t) \\ \dot{\tilde{A}}(t) \end{pmatrix} = \begin{pmatrix} \rho - n - (1-\sigma)x & -\frac{\hat{c}^*}{\hat{k}^*} & \frac{1}{\hat{k}^*}F\left(\hat{k}^*, 1\right) \\ \hat{k}^*\sigma^{-1}F_{11}\left(\hat{k}^*, 1\right) & 0 & \sigma^{-1}F_1\left(\hat{k}^*, 1\right) \\ 0 & 0 & -\phi \end{pmatrix} \begin{pmatrix} \tilde{k}(t) \\ \tilde{c}(t) \\ \tilde{A}(t) \end{pmatrix}$$

ou de façon équivalente,

$$\begin{pmatrix} \dot{\tilde{k}}(t) \\ \dot{\tilde{c}}(t) \\ \dot{\tilde{A}}(t) \end{pmatrix} = \begin{pmatrix} r^* - n - x & -\frac{r^*+\delta}{\alpha} + \delta + n + x & \frac{r^*+\delta}{\alpha} \\ -\frac{1-\alpha}{\sigma\varepsilon}(r^* + \delta) & 0 & \frac{r^*+\delta}{\sigma} \\ 0 & 0 & -\phi \end{pmatrix} \begin{pmatrix} \tilde{k}(t) \\ \tilde{c}(t) \\ \tilde{A}(t) \end{pmatrix} \tag{1.40}$$

en notant que $\frac{\hat{c}^*}{\hat{k}^*} = \frac{\hat{y}^*}{\hat{k}^*} - \delta - n - x = \frac{r^*+\delta}{\alpha} - \delta - n - x$ et avec r^* le taux d'intérêt réel à l'état stationnaire. Le polynôme caractéristique associé à la dynamique linéarisée peut s'écrire de la façon suivante,

$$\chi(\kappa) = (\kappa + \phi)P(\kappa)$$

avec

$$P(\kappa) = \kappa^2 - \kappa\xi + \hat{c}^*\sigma^{-1}F_{11}\left(\hat{k}^*, 1\right)$$

et ξ défini dans la section [2.2]. On note $\kappa_2 = -\phi$ la racine négative liée à la convergence technologique. On montre que les racines du polynôme $P(\kappa)$ sont de signes opposés, $\kappa_0 > 0$ et $\kappa_1 < 0$. En effet, on a,

$$\kappa_0 + \kappa_1 = \rho - n - (1-\sigma)x > 0 \tag{1.41}$$

$$\kappa_0\kappa_1 = \hat{c}^*\sigma^{-1}F_{11}\left(\hat{k}^*, 1\right) < 0 \tag{1.42}$$

L'inégalité (1.41) est une condition nécessaire et suffisante pour que la condition de transversalité soit respectée. L'inégalité (1.42) vient de la convexité de l'ensemble de production. La valeur propre κ_1 correspond donc à la vitesse de convergence néo-classique liée à l'accumulation du capital. Le discriminant associé au polynôme du second degré est donné par,

$$\Delta = \xi^2 - 4\hat{c}^*\sigma^{-1}F_{11}\left(\hat{k}^*, 1\right) > 0$$

En adoptant les notations suivantes pour la part du capital dans la valeur ajoutée et l'élasticité capital–travail,

$$\alpha(t) = \frac{\hat{k}(t)F_1\left(\hat{k}(t),1\right)}{F\left(\hat{k}(t),1\right)}$$

$$\varepsilon(t) = -\frac{F_1\left(\hat{k}(t),1\right)\left[F\left(\hat{k}(t),1\right) - \hat{k}(t)F_1\left(\hat{k}(t),1\right)\right]}{\hat{k}(t)F\left(\hat{k}(t),1\right)F_{11}\left(\hat{k}(t),1\right)}$$

On peut alors réécrire le déterminant de façon équivalente,

$$\Delta = \xi^2 + 4\frac{\hat{c}^*}{\hat{k}^*}\frac{1-\alpha}{\sigma\varepsilon}\left(r^* + \delta\right)$$

où r^* est le taux d'intérêt à l'état stationnaire. En rappelant que le ratio consommation – capital vérifie l'égalité suivante,

$$\frac{\hat{c}^*}{\hat{k}^*} = \frac{r^* + \delta}{\alpha} - (\delta + n + x)$$

On a finalement,

$$\kappa_1 = \frac{\xi}{2} - \frac{1}{2}\sqrt{\xi^2 + \frac{4(1-\alpha)}{\sigma\varepsilon}(\rho + \sigma x + \delta)\left[\frac{\rho + \sigma x + \delta}{\alpha} - \delta - x - n\right]}$$

La vitesse de convergence néo-classique est donnée par $\mu = -\kappa_1$. Notons qu'elle correspond à la vitesse de convergence que nous obtiendrions dans un modèle de croissance optimale sans dynamique de la productivité.
Pour caractériser la dynamique du capital et du produit autour de l'état stationnaire il nous reste à décrire l'espace propre associé au modèle linéarisé.

Vecteur propre associé à la dynamique de productivité

Commençons par calculer le vecteur propre associé à la dynamique de productivité. On note h_2 ce vecteur, il est défini par,

$$\begin{pmatrix} \rho - n - (1-\sigma)x + \phi & -\frac{\rho+\sigma x+\delta}{\alpha} + \delta + n + x & \frac{\rho+\sigma x+\delta}{\alpha} \\ -\frac{1-\alpha}{\sigma\varepsilon}\left(\rho + \sigma x + \delta\right) & \phi & \frac{\rho+\sigma x+\delta}{\sigma} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} h_{21} \\ h_{22} \\ h_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Après quelques calculs, on obtient des expressions de h_{23} et h_{22} en fonction de h_{21},

$$h_{23} = -\frac{1-\alpha}{\varepsilon} \frac{\left(1 + \frac{\phi-\delta-n-x}{r^*+\delta}\right)\frac{\phi\varepsilon}{1-\alpha} + \frac{1}{\sigma}\left(\delta + n + x - \frac{r^*+\delta}{\alpha}\right)}{\frac{\phi}{\alpha} - \frac{1}{\sigma}\left(\delta + n + x - \frac{r^*+\delta}{\alpha}\right)} h_{21}$$

$$h_{22} = \frac{r^*+\delta}{\sigma\phi}\frac{1-\alpha}{\varepsilon}\left(1 + \frac{\left(1 + \frac{\phi-\delta-n-x}{r^*+\delta}\right)\frac{\phi\varepsilon}{1-\alpha} + \frac{1}{\sigma}\left(\delta + n + x - \frac{r^*+\delta}{\alpha}\right)}{\frac{\phi}{\alpha} - \frac{1}{\sigma}\left(\delta + n + x - \frac{r^*+\delta}{\alpha}\right)}\right) h_{21}$$

Vecteur propre associé au mécanisme de convergence néo-classique

On cherche le vecteur propre h_1 dans $\mathbb{R}^3$ défini par,

$$\begin{pmatrix} \rho - n - (1-\sigma)x - \kappa_1 & -\frac{\rho+\sigma x+\delta}{\alpha} + \delta + n + x & \frac{\rho+\sigma x+\delta}{\alpha} \\ -\frac{1-\alpha}{\sigma\varepsilon}\left(\rho + \sigma x + \delta\right) & -\kappa_1 & \frac{\rho+\sigma x+\delta}{\sigma} \\ 0 & 0 & -\phi - \kappa_1 \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Clairement on a $h_{13} = 0$. Le problème se réduit donc a

$$\begin{pmatrix} r^* - n - x - \kappa_1 & -\frac{r^*+\delta}{\alpha} + \delta + n + x \\ -\frac{1-\alpha}{\sigma\varepsilon}\left(r^* + \delta\right) & -\kappa_1 \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

on obtient une expression de h_{12} en fonction de h_{11},

$$h_{12} = -\frac{1-\alpha}{\kappa_1\sigma\varepsilon}(r^* + \delta)h_{11}$$

Vecteur propre associé au bras instable de la dynamique

On cherche le vecteur propre h_0 dans $\mathbb{R}^3$ défini par,

$$\begin{pmatrix} \rho - n - (1-\sigma)x - \kappa_0 & -\frac{\rho+\sigma x+\delta}{\alpha} + \delta + n + x & \frac{\rho+\sigma x+\delta}{\alpha} \\ -\frac{1-\alpha}{\sigma\varepsilon}\left(\rho + \sigma x + \delta\right) & -\kappa_0 & \frac{\rho+\sigma x+\delta}{\sigma} \\ 0 & 0 & -\phi - \kappa_0 \end{pmatrix} \begin{pmatrix} h_{01} \\ h_{02} \\ h_{03} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

De la même façon que dans le cas précédant, on obtient

$$h_{02} = -\frac{1-\alpha}{\kappa_0 \sigma \varepsilon}(r^* + \delta) h_{01}$$

et

$$h_{02} = 0$$

La dynamique linéarisée autour de l'état stationnaire s'écrit alors de la façon suivante,

$$X(t) = a_0 h_0 e^{\kappa_0 t} + a_1 h_1 e^{\kappa_1 t} + a_2 h_2 e^{\kappa_2 t} \ \forall t \geq 0$$

où $(a_0, a_1, a_2) \in \mathbb{R}^3$ et $X(t) = (\tilde{k}(t), \tilde{c}(t), \tilde{A}(t))'$. Pour que l'on ne viole pas la condition de transversalité il faut que $X(0)$ appartienne à l'espace engendré par les vecteurs colonnes h_1 et h_2 (associés aux valeurs propres négatives de la dynamique). On doit donc poser $a_0 = 0$ et

$$X(0) = a_1 h_1 + a_2 h_2$$

On peut alors, étant données les conditions initiales sur les deux variables d'état, définir les coefficients a_1 et a_2. On a les deux égalités suivantes,

$$\tilde{k}(0) = a_1 + a_2$$

$$\tilde{A}(0) = -\frac{1-\alpha}{\varepsilon} \frac{\left(1 + \frac{\phi - \delta - n - x}{r^* + \delta}\right) \frac{\phi \varepsilon}{1-\alpha} + \frac{1}{\sigma}\left(\delta + n + x - \frac{r^* + \delta}{\alpha}\right)}{\frac{\phi}{\alpha} - \frac{1}{\sigma}\left(\delta + n + x - \frac{r^* + \delta}{\alpha}\right)} a_2$$

On a donc,

$$a_2 = -\frac{\varepsilon}{1-\alpha}\gamma^{-1}\tilde{A}(0)$$

et

$$a_1 = \tilde{k}(0) + \frac{\varepsilon}{1-\alpha}\gamma^{-1}\tilde{A}(0)$$

où γ est donné par,

$$\gamma = \frac{\sigma\varepsilon}{1-\alpha} \frac{(\xi + \phi)\phi - \frac{1-\alpha}{\sigma\varepsilon}(\rho + \sigma x + \delta)\left(\frac{\rho + \sigma x + \delta}{\alpha} - \delta - n - x\right)}{(\rho + \sigma x + \delta)\left(\frac{\rho + \sigma(x + \phi) + \delta}{\alpha} - \delta - n - x\right)}$$

On montre que le signe de γ dépend des positions relatives des deux vitesses de convergence. Nous savons que $\kappa_0 + \kappa_1 = \xi > 0$ par la condition de transversalité, ainsi $(\xi+\phi)\phi > 0$. Par ailleurs,

$$\kappa_0\kappa_1 = -\frac{1-\alpha}{\sigma\varepsilon}\frac{\rho+\sigma x+\delta}{\sigma}\left(\frac{\rho+\sigma x+\delta}{\alpha}-\delta-n-x\right)$$

Puisque l'état stationnaire de ce modèle est déterminé, on sait que $\kappa_0\kappa_1 < 0$ et donc que l'inégalité suivante doit être satisfaite,

$$\frac{\rho+\sigma x+\delta}{\alpha} > \delta+n+x$$

Cette inégalité entraîne que le dénominateur dans l'expression de γ est toujours positif. Le numérateur est négatif si et seulement si le terme positif $(\xi+\phi)\phi$ n'est pas trop important. On peut trouver une valeur de ϕ telle que γ est nul, en résolvant l'équation du second degré en ϕ,

$$\mathcal{E}(\phi) = \phi^2 + \xi\phi - \frac{1-\alpha}{\sigma\varepsilon}(r^*+\delta)\frac{\hat{c}^*}{\hat{k}^*} = 0$$

En notant que $\mathcal{E}(0) < 0$ et que le coefficient associé à ϕ^2 est positif, on voit que cette équation possède deux racines réelles de signes opposés. On choisit la racine positive, puisque ϕ est la vitesse de rattrapage technologique (positive par hypothèse). Ainsi, en notant ϕ^* la valeur de ϕ qui annule γ, on a,

$$\begin{aligned}\phi^* &= \frac{-\xi+\sqrt{\xi^2-\frac{4(1-\alpha)}{\sigma\varepsilon}(\rho+\sigma x+\delta)\left[\frac{\rho+\sigma x+\delta}{\alpha}-\delta-n-x\right]}}{2}\\ \Leftrightarrow \phi^* &= \mu\end{aligned}$$

Le paramètre γ est négatif si et seulement si $\phi < \mu$. Ainsi, le terme devant l'écart de productivité initial dans l'équation (1.4), décrivant la dynamique du capital par tête efficace en déviation à l'état stationnaire, est toujours positif. ■

Convergence conditionnelle. Pour envisager la convergence conditionnelle, on suppose que le produit de l'économie θ s'écrit maintenant de la façon suivante,

$$Y(\theta,t) = \tau(\theta)A(\theta,t)F(K(\theta,t),\mathcal{T}(t)L(t))$$

Les économies sont hétérogènes en termes des conditions initiales $A(\theta, t)$ et $K(\theta, t)$, mais aussi en termes du paramètre invariant dans le temps $\tau(\theta)$. C'est l'hétérogénéité sur ce paramètre qui va entraîner l'hétérogénéité des états stationnaires. Ce paramètre peut, par exemple, représenter une taxe sur le produit spécifique à chaque économie. Nous pourrions aussi introduire une hétérogénéité sur le niveau initial de la population. Cette hétérogénéité n'a pas d'impact sur les valeurs propres de la dynamique (*ie* l'expression du jacobien associé à la dynamique linéarisée demeure identique) ainsi la vitesse de convergence n'est pas affectée. Les résultats qui suivent sont robuste à une convergence conditionnelle. Il suffit, par le théorème de Frisch-Waugh, de régresser au préalable les variables explicatives et la variable expliquée sur les variables qui permettent de contrôler de l'hétérogénéité des états stationnaires.

4.2 Le cas sans dynamique de productivité

La proposition envisage le cas où il n'y a pas de dynamique de productivité et montre que la régression à la Barro donne des estimations qui convergent vers la vraie valeur des paramètres (essentiellement la vitesse de convergence néo-classique).

Proposition 4.1. *S'il n'y a pas de dynamique de productivité, c'est à dire si* $\mathbb{E}[\theta_2] = \mathbb{V}[\theta_2] = 0$*, alors on obtient*

$$\begin{aligned} b(T) &= \mu && (1.43) \\ a(T) &= x + \frac{1 - e^{-\mu T}}{T} && (1.44) \end{aligned}$$

L'estimateur de la vitesse de convergence ne dépend pas de la dimension chronologique considérée ni de l'état initial du monde (espérance ou variance de la condition initiale).

Preuve On procède comme pour les preuves des propositions 2.1 et 2.2, voir plus bas. ∎

Dans ce cas, une régression à la Barro est capable d'identifier la vitesse de convergence néo-classique. Par la suite, dès lors que le modèle est augmenté d'une variable d'état, on montre que ce résultat favorable à cette méthodologie ne tient plus. Si ce type de régression peut, sous certaines conditions, estimer de façon convergente une vitesse de convergence celle-ci sera sans rapport a priori avec le mécanisme de convergence néo-classique.

4.3 Preuve de la proposition 2.1

Le lemme suivant exprime l'estimateur $\hat{b}_\infty(T)$ et servira dans les preuves des propositions 2.1 et 2.2.

Lemme 4.2. *La limite en probabilité de la vitesse de convergence obtenue à l'aide d'une régression à la Barro est de la forme suivante,*

$$\hat{b}_\infty(T) = \phi - \frac{1}{T}\log B(T)$$

avec

$$B(T) = \frac{e^{-(\mu-\phi)T}\{\alpha^2 a + \alpha c\} + b + \alpha c + \left(e^{-(\mu-\phi)T} - 1\right)\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\{b + \alpha c\}}{\alpha^2 a + b + 2\alpha c}$$

où $a = \mathbb{V}[\theta_1]$, $b = \mathbb{V}[\theta_2]$ et $c = \mathbb{C}ov(\theta_1, \theta_2)$.

Preuve du lemme 4.2 A partir des conditions nécessaires (1.10) et (1.11) on obtient l'expression suivante pour définir $\hat{b}_\infty(T)$,

$$\frac{1 - e^{-\hat{b}_\infty(T)}}{T} = -\frac{\int_\Theta g_T(\theta)\log y(\theta,0)d\Phi(\theta) - \left(\int_\Theta g_T(\theta)d\Phi(\theta)\right)\left(\int_\Theta \log y(\theta,0)d\Phi(\theta)\right)}{\int_\Theta \log^2 y(\theta,0)d\Phi(\theta) - \left(\int_\Theta \log y(\theta,0)d\Phi(\theta)\right)^2}$$

En utilisant le processus générateur des données, on a les expressions suivantes,

$$\begin{aligned}\int_\Theta g_T(\theta)d\Phi(\theta) = x - \frac{1-e^{-\mu T}}{T}\alpha\mathbb{E}[\theta_1] \quad &- \quad \frac{1-e^{-\phi T}}{T}\mathbb{E}[\theta_2] \\ &+ \quad \frac{e^{-\mu T} - e^{-\phi T}}{T}\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\mathbb{E}[\theta_2]\end{aligned}$$

On a $\log y(\theta,0) \equiv \breve{y}(\theta) = \tilde{y}(\theta,0) + \breve{y}^* = \tilde{k}(\theta,0) + \widetilde{A}(\theta,0) + \breve{y}^*$.

$$\begin{aligned}
\int_\Theta \log y(\theta,0) d\Phi(\theta) &= \alpha\mathbb{E}[\theta_1] + \mathbb{E}[\theta_2] + \breve{y}^* \\
\int_\Theta g_T(\theta)\breve{y}(\theta) d\Phi(\theta) &= x\left(\breve{y}^* + \alpha\mathbb{E}[\theta_1]\mathbb{E}[\theta_2]\right) \\
&- \frac{1-e^{-\mu T}}{T}\left(\alpha^2\mathbb{E}[\theta_1^2] + \alpha\mathbb{E}[\theta_1\theta_2] + \alpha\breve{y}^*\mathbb{E}[\theta_1]\right) \\
&- \frac{1-e^{-\phi T}}{T}\left(\mathbb{E}[\theta_2^2] + \alpha\mathbb{E}[\theta_1\theta_2] + \breve{y}^*\mathbb{E}[\theta_2]\right) \\
&+ \frac{e^{-\mu T}-e^{-\phi T}}{T}\frac{\alpha\varepsilon\gamma^{-1}}{1-\alpha}\left(\mathbb{E}[\theta_2^2] + \alpha\mathbb{E}[\theta_1\theta_2] + \breve{y}^*\mathbb{E}[\theta_2]\right)
\end{aligned}$$

Notons enfin que le terme au dénominateur de l'expression définissant $\hat{b}_\infty(T)$ correspond à la variance de $\breve{y}(\theta) = \tilde{y}(\theta,0) + \breve{y}^*$, c'est à dire à

$$\mathbb{V}[\breve{y}(\theta)] = \alpha^2\mathbb{V}[\theta_1] + \mathbb{V}[\theta_2] + 2\alpha\mathbb{C}ov(\theta_1,\theta_2)$$

puisque $\breve{y}^*$ est déterministe. En substituant ces expressions dans l'expression de $\hat{b}_\infty(T)$ puis en factorisant $e^{\phi T}$ on trouve bien la formulation donnée dans le lemme 4.2. ■

Preuve de la proposition 2.1 On utilise le lemme 4.2 en posant $\mathbb{C}ov(\theta_1,\theta_2) = 0$. On obtient l'expression donnée dans la proposition 2.1 pour la limite en probabilité de l'estimateur de la vitesse de convergence. Elle dépend de la distribution de l'état initial du monde (via les variances de θ_1 et θ_2). Clairement l'estimateur de la vitesse de convergence tend vers ϕ lorsque $T \to \infty$, ce qui prouve le point ***(i)***.

On montre que pour une paramétrisation donnée du modèle, $\hat{b}_\infty(T) > \phi$ ou $\hat{b}_\infty(T) < \phi$ pour tout T, c'est à dire que la fonction $\hat{b}_\infty(T)$ ne peut traverser ϕ. En effet on sait que $B(0) = 1$ et qu'asymptotiquement on a,

$$\lim_{T\to\infty} B(T) = \frac{1 - \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}}{\alpha^2\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} + 1} \equiv B^*$$

$B^* > 1$ si et seulement si l'inégalité suivante est vérifiée,

$$\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} < -\frac{\varepsilon\gamma^{-1}}{\alpha(1-\alpha)} \tag{1.45}$$

Il s'agit aussi d'une condition nécessaire et suffisante pour que la dérivée,

$$\frac{\partial B(T)}{\partial T} = -(\mu - \phi)e^{-(\mu-\phi)T}\frac{\alpha^2\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} + \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}}{\alpha^2\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]} + 1}$$

soit positive. Si l'inégalité (1.45) est vérifiée alors $B(T) > 1$ pour tout $T > 0$ et donc $b(T) < \phi$ pout tout T vérifiant $0 < T < \infty$, sinon $b(T) > \phi$.

Montrons que la vitesse de convergence obtenue avec une régression à la Barro est encadrée par la μ et la fonction $\underline{b}(T)$. On note que, pour tout horizon T, la vitesse de convergence $\hat{b}_\infty(T)$ est une fonction croissante du ratio $\nu \equiv \mathbb{V}[\theta_1]/\mathbb{V}[\theta_2]$. On obtient facilement obtient le comportement limite de $\hat{b}_\infty(T)$ en fonction de ν,

$$\begin{aligned}
\lim_{\nu\to\infty} b(T) &= \mu \equiv \overline{b}(T) \\
\lim_{\nu\to 0} b(T) &= \phi - \frac{1}{T}\log\left(1 + \left(e^{-(\mu-\phi)T} - 1\right)\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right) \equiv \underline{b}(T)
\end{aligned}$$

La vitesse de convergence est donc bornée supérieurement par $\overline{b}(T)$, la vitesse de convergence néo-classique, et inférieurement par une fonction de $\underline{b}(T)$. La borne inférieure des valeurs possibles de $\hat{b}_\infty(T)$ correspond au cas limite où l'hétérogénéité du produit par tête s'explique exclusivement par la productivité. Même si la variance du capital par tête en déviation à l'état stationnaire est initialement nulle, on n'a pas $b(T) = \phi$ car l'inefficacité technologique des économies induit une dynamique du capital par tête qui s'éloigne de l'état stationnaire. Ceci explique pourquoi dans ce cas on mesure une vitesse de convergence inférieure à ϕ alors même que seule la productivité est à l'origine de la dynamique. En conséquence la vitesse de convergence est toujours inférieure à μ, ce qui prouve le point ***(ii)***.

Pour montrer le point ***(iii)***, il suffit de considérer la dérivée suivante,

$$\frac{\partial \hat{b}_\infty(T)}{\partial T} = \frac{1}{T^2}\left(\log B(T) - T\frac{\partial B(T)/\partial T}{B(T)}\right)$$

Clairement, si l'inégalité (1.45) est vérifiée alors il existe un rang T^* tel que la dernière expression est positive pour tout $T > T^*$. Dans ce cas $\hat{b}_\infty(T)$ tend vers ϕ par dessous, c'est à dire,

$$\hat{b}_\infty(T) \underset{T\to\infty}{\nearrow} \phi$$

Si l'inégalité (1.45) n'est pas vérifiée $\hat{b}_\infty(T)$ tend vers ϕ par dessus. ∎

4.4 Preuve de la proposition 2.2 et plus

Montrons que la vitesse de convergence est définie pour tout horizon T si et seulement si $\mathbb{C}ov(\alpha\theta_1, \theta_2) + \mathbb{V}[\theta_2] > 0$. On a nécessairement $\alpha^2\mathbb{V}[\theta_1] + \mathbb{V}[\theta_2] + 2\alpha\mathbb{C}ov(\theta_1, \theta_2) > 0$ pour assurer que la variance du produit initial par travailleur efficace est strictement positive. La vitesse de convergence est définie si et seulement si l'expression sous le logarithme est strictement positive, c'est à dire si et seulement si on a,

$$\begin{aligned}\Gamma(T) = e^{-(\mu-\phi)T}(\alpha^2\mathbb{V}[\theta_1] + \chi) \quad &+ \quad \mathbb{V}[\theta_2] + \chi \\ &+ \quad \left(e^{-(\mu-\phi)T} - 1\right)\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}(\mathbb{V}[\theta_2] + \chi) > 0\end{aligned}$$

On a

$$\Gamma(0) = \mathbb{V}[\tilde{y}(\theta, 0)] > 0$$

et

$$\lim_{T\to\infty} \Gamma(T) = (\mathbb{V}[\theta_2] + \chi)\left(1 - \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)$$

Cette dernière expression est strictement positive si et seulement si $\chi > -\mathbb{V}[\theta_2]$. En notant que la dérivée de $\Gamma(T)$ par rapport à T est donnée par

$$\frac{\partial\Gamma(T)}{\partial T} = -(\mu-\phi)e^{-(\mu-\phi)T}\left[\alpha^2\mathbb{V}[\theta_1] + \chi + \frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}(\mathbb{V}[\theta_2] + \chi)\right]$$

Puisque, pour χ donné, la fonction $\Gamma(T)$ apparaît monotone, $\mathbb{C}ov(\alpha\theta_1, \theta_2) > -\mathbb{V}[\theta_2]$ est une condition nécessaire et suffisante pour définir la vitesse de convergence pour tout T. On peut exprimer cette condition de façon équivalente en considérant la corrélation entre θ_1 et θ_2,

$$r(\theta_1, \theta_2)\sqrt{\frac{\mathbb{V}[\theta_1]}{\mathbb{V}[\theta_2]}} > -\frac{1}{\alpha}$$

Ainsi on voit que cette restriction *ne mord pas* si la variance du capital n'est pas trop grande par rapport à la variance de l'efficience.

En reprenant l'expression donnée dans le lemme 4.2 on obtient immédiatement l'expression de l'estimateur de la vitesse de convergence donné dans la proposition 2.2. Encore une fois la vitesse de convergence estimée dépend de l'état initial du monde. Clairement, l'estimateur de la vitesse de convergence tend vers ϕ lorsque T tend vers l'infini. Pour le reste on procède comme dans la preuve de la proposition 2.1. Pour montrer que la vitesse de convergence mesurée à l'aide d'une régression à la Barro est encadrée de la même façon par $\overline{b}(T) = \mu$ et $\underline{b}(T)$, il faut noter que l'on a nécessairement $\mathbb{C}ov(\theta_1, \theta_2) \sim O(\sqrt{\mathbb{V}[\theta_1]\mathbb{V}[\theta_2]})$ afin de ne pas violer l'inégalité de Cauchy-Schwartz. ■

On peut montrer que la vitesse de convergence mesurée par une régression à la Barro est une fonction croissante de la covariance entre θ_1 et θ_2 si et seulement si $\alpha^2\mathbb{V}[\theta_1] < \mathbb{V}[\theta_2]$. On montre facilement que $\hat{b}^c_\infty(T)$ est une fonction croissante de

$$K(\chi) = \frac{\alpha^2\mathbb{V}[\theta_1] + \chi}{\mathbb{V}[\theta_2] + \chi}$$

en observant que $K(\chi)$ est une fonction croissante de χ si et seulement si $\alpha^2\mathbb{V}[\theta_1] < \mathbb{V}[\theta_2]$, on obtient le résultat annoncé.

4.5 Preuve de la proposition 2.3

Après quelques calculs sans grand intérêt on obtient l'équivalence suivante,

$$\hat{b}^c_\infty(T) > b^*(T)$$

$$\Leftrightarrow [K(\chi)-1]\left[e^{-(\mu-\phi)T}-1\right]\left[1-\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right]<0$$

Avec

$$K(\chi)=\frac{\alpha^2\mathbb{V}[\theta_1]+\chi}{\mathbb{V}[\theta_2]+\chi}$$

En rappelant que par hypothèse $\mathbb{V}[\theta_2]+\chi>0$, une condition nécessaire et suffisante pour que $\hat{b}^c_\infty(T)>b^*(T)$ est $\alpha^2\mathbb{V}[\theta_1]>\mathbb{V}[\theta_2]$. ■

4.6 Preuve du corollaire 2.4

D'après la proposition 2.2 on a $\hat{b}_\infty(T)=\phi-T^{-1}\log B^c(T)$ où

$$(K(\chi)+1)\,B^c(T)=e^{-(\mu-\phi)T}\left(K(\chi)+\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)+1-\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}$$

et

$$K(\chi)=\frac{\alpha^2\mathbb{V}[\theta_1]+\chi}{\mathbb{V}[\theta_2]+\chi}$$

On peut redéfinir la vitesse de convergence en fonction de la corrélation $r=\chi/\sqrt{\mathbb{V}[\theta_1]\mathbb{V}[\theta_2]}$ plutôt que la covariance. Il suffit de remplacer $K(\chi)$ par $K(r)$ dans l'expression de $B^c(T)$. On a,

$$K(r)=\frac{r+\alpha\sqrt{\nu}}{r+\frac{\alpha}{\sqrt{\nu}}}$$

Par hypothèse la corrélation est dans l'intervalle $[0,1]$ et on sait que la vitesse de convergence est une fonction monotone de la corrélation. Sachant la vitesse de convergence au sens de Barro β, on peut donc construire un intervalle des valeurs possibles de ν en considérant la réciproque de la définition de la vitesse de convergence pour $r=0$ et $r=1$.
Pour $r=0$, on a $K(0)=\nu$ et donc puisque $B^c(T)=e^{-(\beta-\phi)T}$,

$$(1+\nu)e^{-(\beta-\phi)T}=e^{-(\mu-\phi)T}\left(\nu+\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}\right)+1-\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}$$

d'où

$$\nu^*=\frac{e^{-(\mu-\phi)T}\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}-e^{-(\beta-\phi)T}+1-\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}}{e^{-(\beta-\phi)T}-e^{-(\mu-\phi)T}}$$

En notant que pour tout T on a $\beta < \mu$ dès lors que ν est fini, il vient que ν^* est positif.
Pour $r = 1$, on a

$$K(1) = \frac{1 + \alpha\sqrt{\nu}}{1 + \frac{\alpha}{\sqrt{\nu}}}$$

et donc

$$K(1) = \nu^*$$

Comme $K(1)$ est une fonction continue monotone croissante en ν, il existe une solution notée ν^{**}. En adoptant la notation $b^{**} = \sqrt{\nu^{**}}$, ν^{**} est défini par la racine carrée de la racine positive de,

$$\mathcal{E}(b^{**}) = \alpha b^{**2} + (1 - \nu^*)b^{**} - \alpha\nu^* = 0$$

En remarquant que $\mathcal{E}(0) < 0$, on sait qu'il n'existe qu'une seule solution réelle positive. Elle est donnée par l'équation (1.16) dans le corollaire 2.4. ■

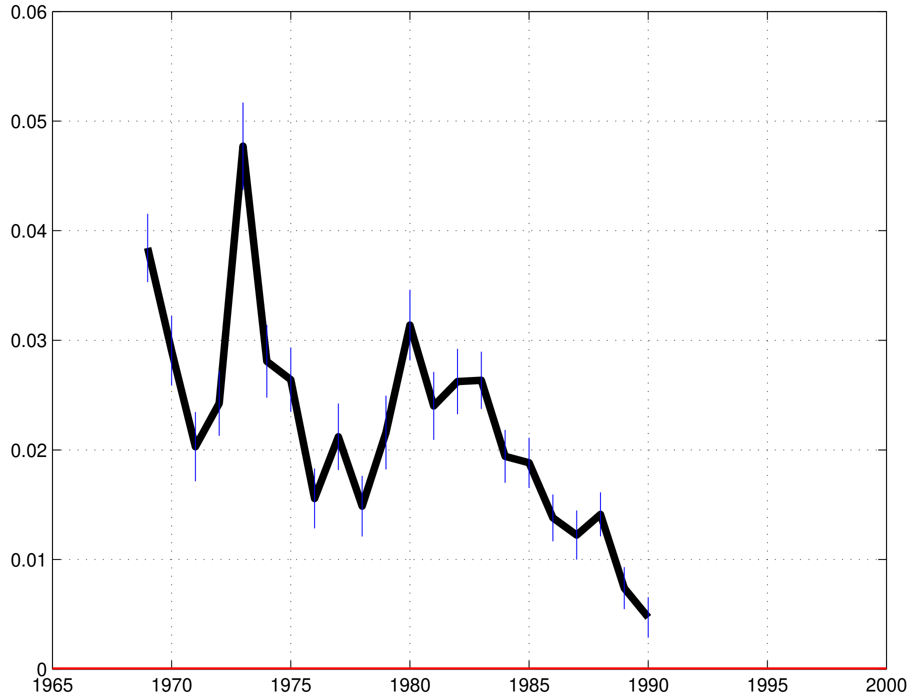

Fig. 1.3 – **Vitesses de convergence pour 3075 comtés américains.** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête sur la condition initiale et des indicatrices régionales. Les barres verticales représentent des intervalles de confiance au seuil de 5%.

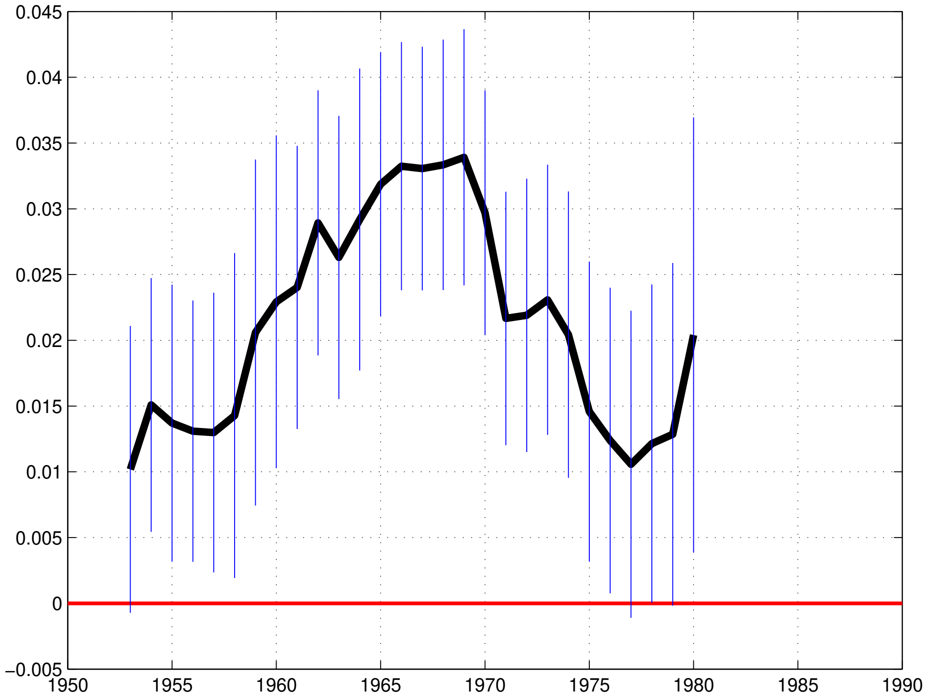

Fig. 1.4 – **Vitesses de convergence pour 27 économies de l'OCDE.** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête sur la condition initiale et une constante. Les barres verticales représentent des intervalles de confiance au seuil de 5%.

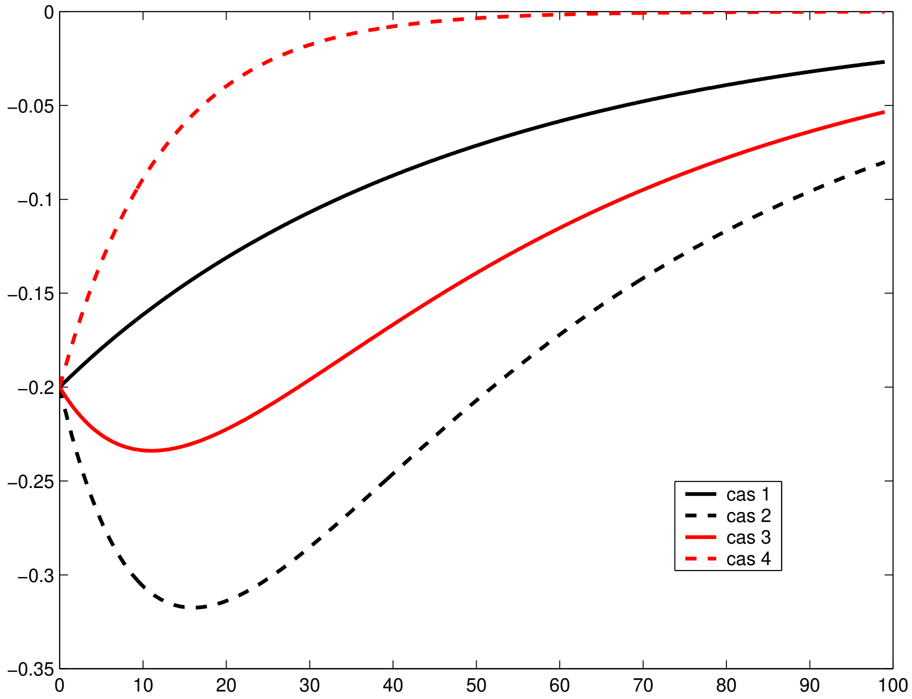

Fig. 1.5 – **Dynamique du produit par travailleur.**

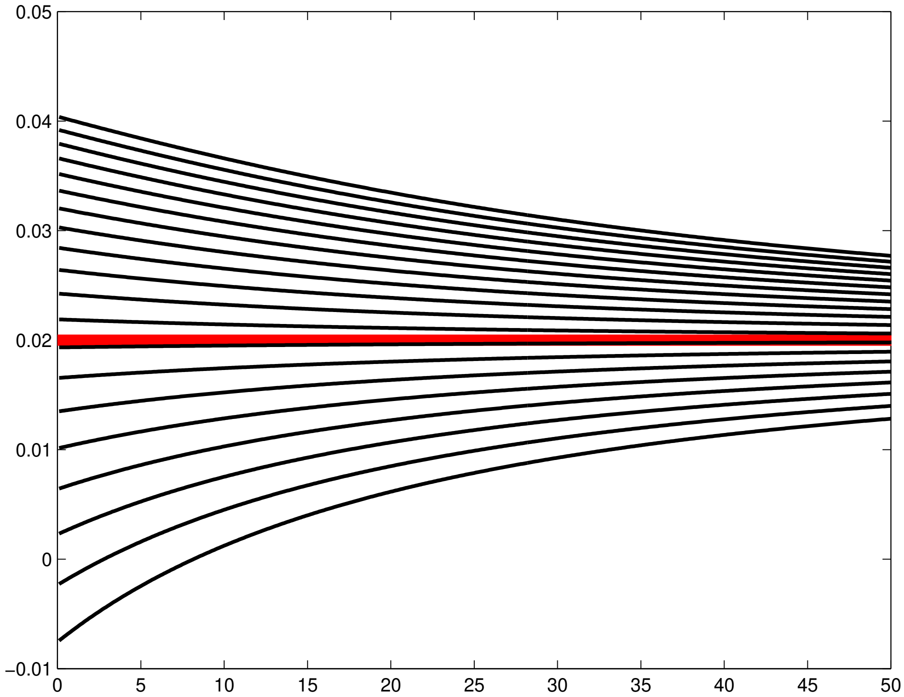

Fig. 1.6 – **Vitesse de convergence & régression à la Barro (I).** Chaque courbe représente la vitesse de convergence que l'on obtient à l'aide d'une régression à la Barro en fonction de l'horizon T. On a posé $\mathbb{V}[\tilde{A}(0)] = 1$ et $\mathbb{V}[\tilde{k}(0)] = 1$ pour la première courbe au sud, $\mathbb{V}[\tilde{k}(0)] = 2$ pour la deuxième courbe en allant vers le nord, $\mathbb{V}[\tilde{k}(0)] = 3$ pour la troisième courbe,...

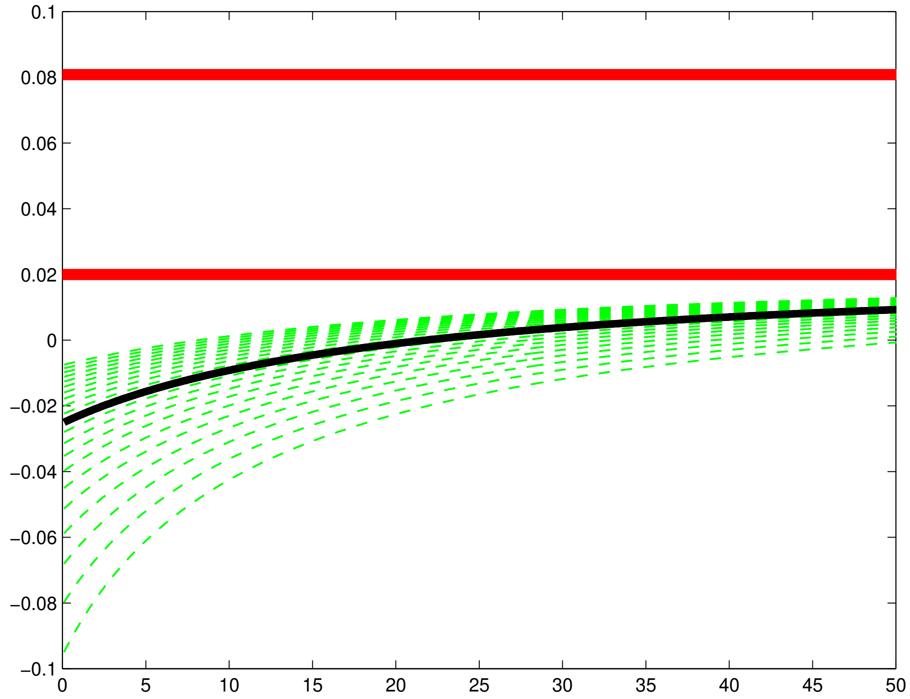

FIG. 1.7 – **Vitesse de convergence & régression à la Barro (II).** Les droites horizontales rouges correspondent aux vitesses de convergence technologique et néo-classique. La courbe noire et pleine correspond à la vitesse de convergence mesurée par une régression à la Barro lorsque le ratio de la variance de θ_1 à la variance de θ_2 est unitaire. Les courbes vertes en tirets correspondent aux vitesses de convergence que nous pourrions alors obtenir en faisant varier la covariance entre θ_1 et θ_2.

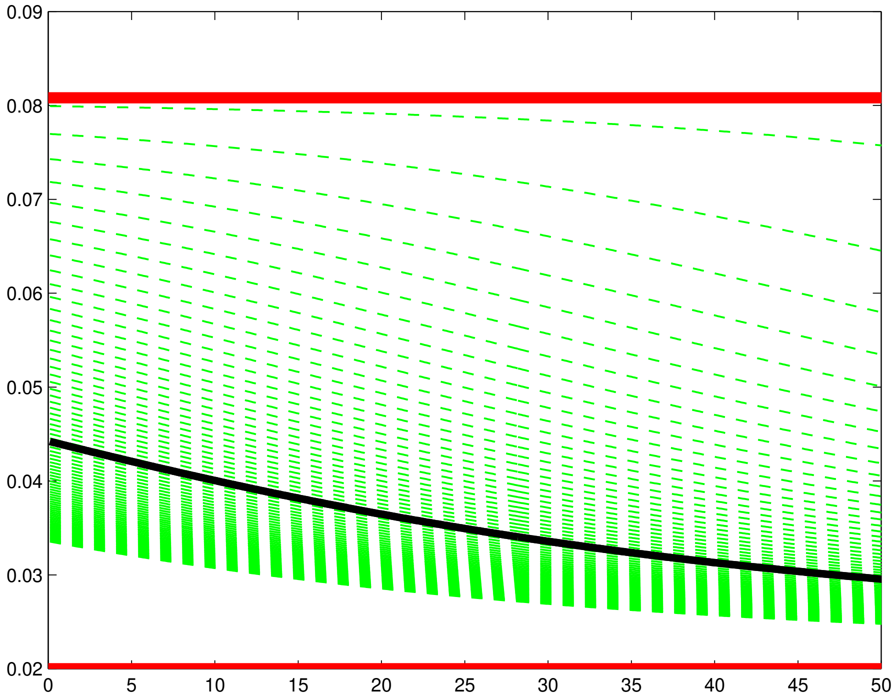

Fig. 1.8 – **Vitesse de convergence & régression à la Barro (III).** Les droites horizontales rouges correspondent aux vitesses de convergence technologique et néo-classique. La courbe noire et pleine correspond à la vitesse de convergence mesurée par une régression à la Barro lorsque le ratio de la variance de θ_1 à la variance de θ_2 est égal à 20. Les courbes vertes en tirets correspondent aux vitesses de convergence que nous pourrions obtenir en faisant varier la covariance entre θ_1 et θ_2.

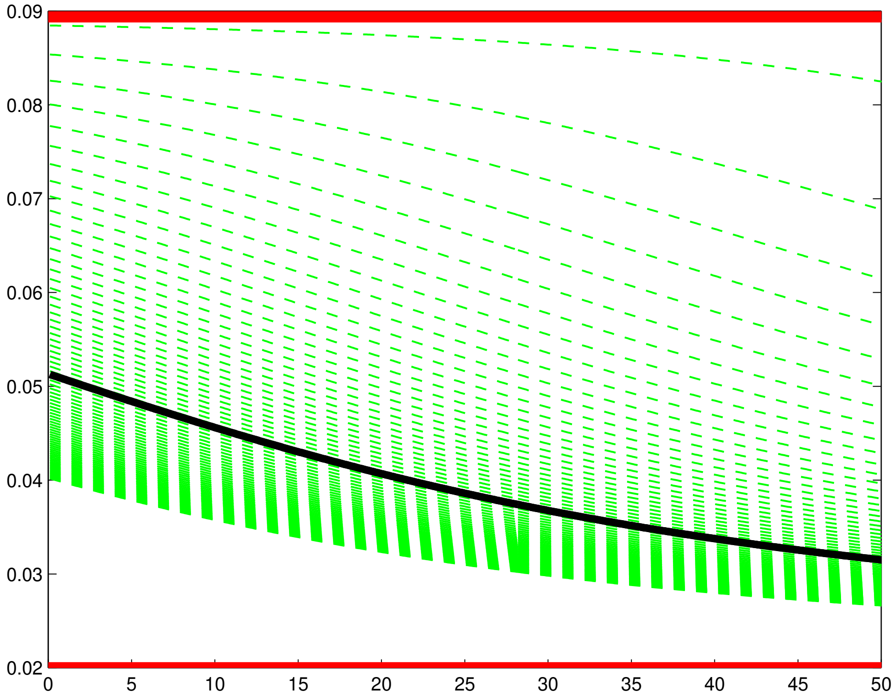

Fig. 1.9 – **Vitesse de convergence & régression à la Barro (IV).** Même graphique que [1.8] avec $\varepsilon = 1$ (*ie*, technologie de production Cobb-Douglas).

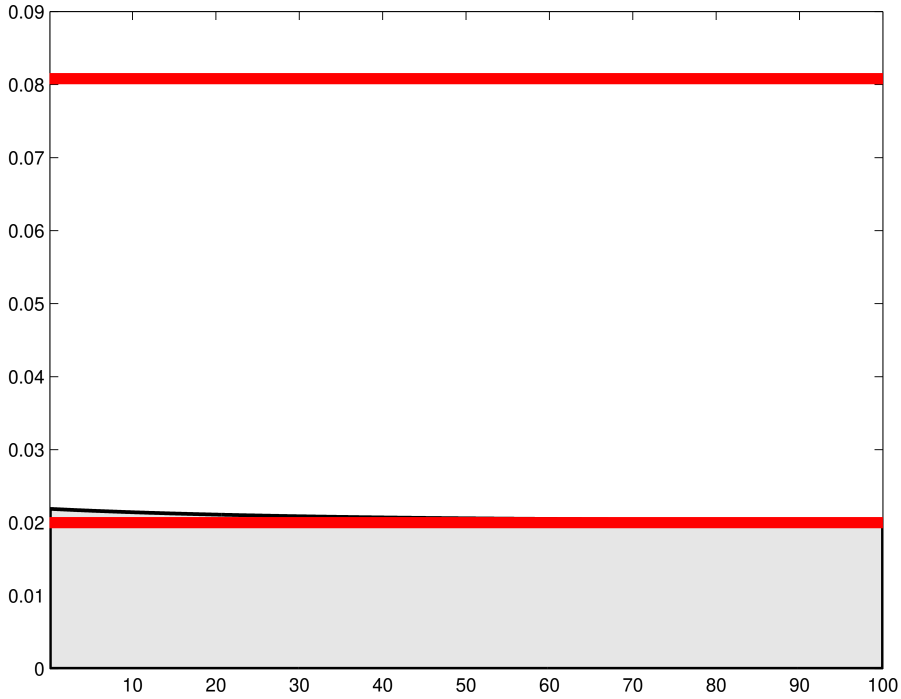

Fig. 1.10 – **Capital versus Technologie.**

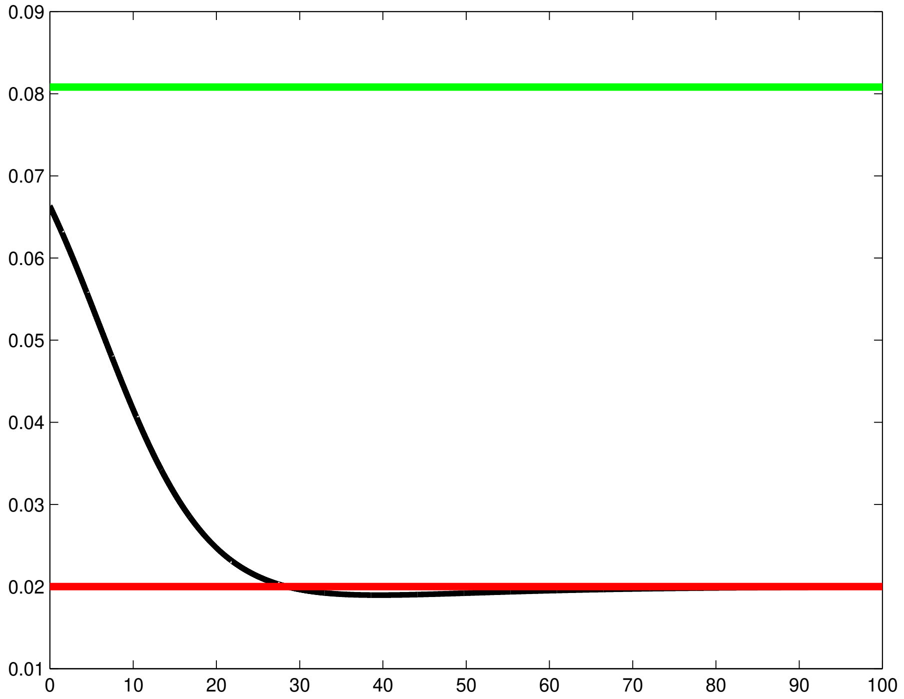

Fig. 1.11 – **Vitesse de convergence & régression à la Barro (V).**

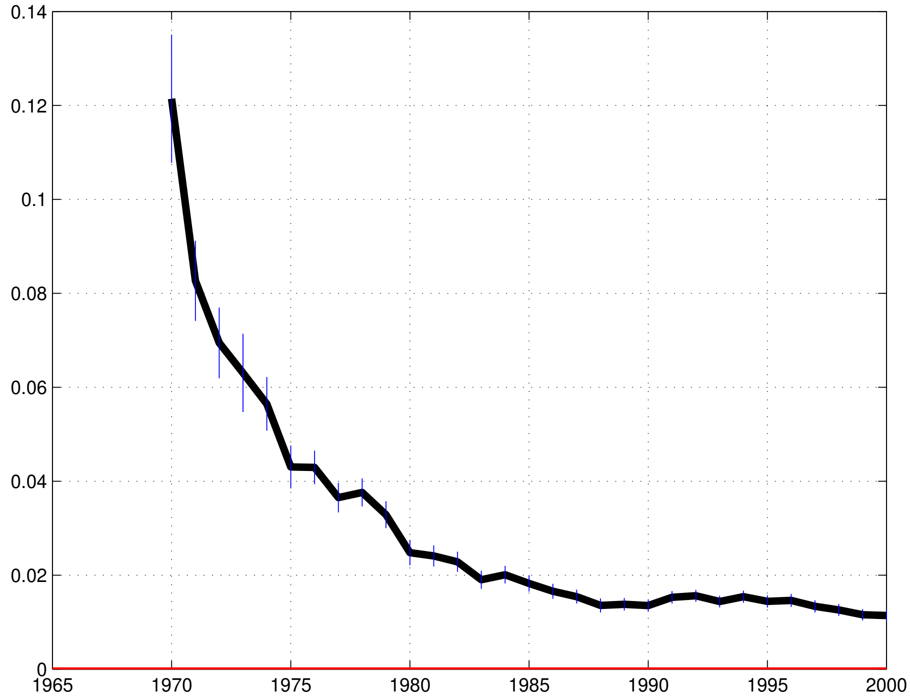

Fig. 1.12 – **Vitesses de convergence pour 3075 comtés américains.** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête sur la condition initiale et des indicatrices régionales. Les barres verticales représentent des intervalles de confiance au seuil de 5%.

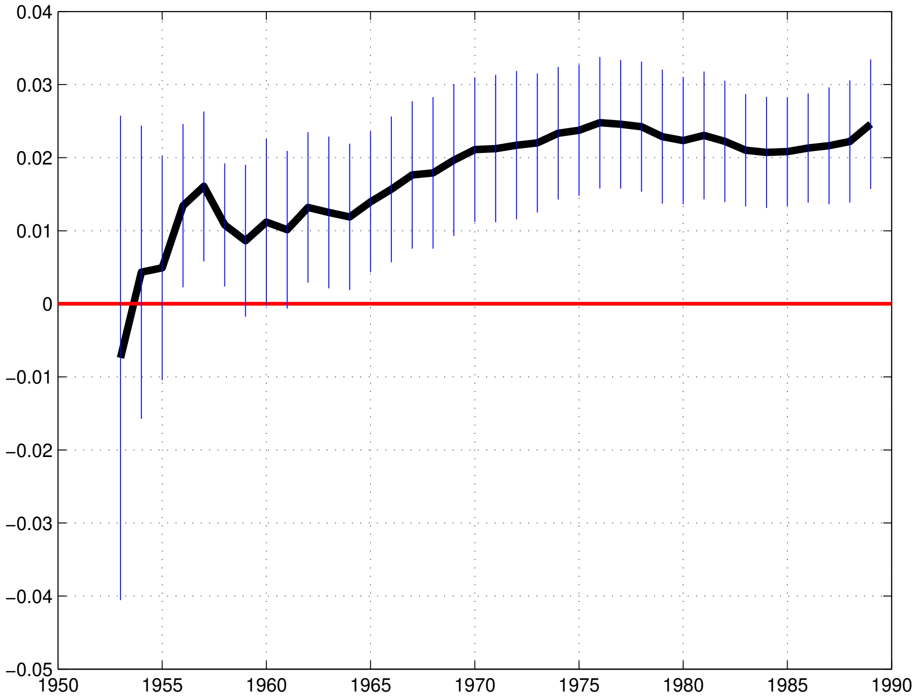

Fig. 1.13 – **Vitesses de convergence pour 27 économies de l'OCDE.** Ces estimations, par les moindres carrés non linéaires, sont obtenues en régressant le taux de croissance moyen du revenu personnel par tête sur la condition initiale et une constante. Les barres verticales représentent des intervalles de confiance au seuil de 5%.

Périodes	$\hat{\beta}$	$\hat{\beta}^*$
1880–1900	0,0224	0,0268
1900–1920	0,0209	0,0269
1920–1930	−0,0122	0,0218
1930–1940	0,0127	0,0141
1940–1950	0,0373	0,0362
1950–1960	0,0202	0,0313
1960–1970	0,0135	0,0194
1970–1980	0,0119	0,0139
1980–1988	−0,0005	0,0196
χ^2	32,1	13,9

TAB. 1.1 – **Estimations de la vitesse de convergence entre les états américains.** La dernière ligne donne la statistique du ratio de vraisemblance pour tester l'égalité des vitesses de convergence (la valeur critique au seuil de 5% est 15,5). Dans la deuxième colonne, la vitesse de convergence est obtenue en régressant le taux de croissance moyen sur le niveau initial et des indicatrices régionales. Dans la troisième colonne, la vitesse de convergence est estimée en régressant le taux de croissance moyen sur la condition initiale, des indicatrices régionales et une variable contrôlant des chocs agrégés qui affectent les états. **Sources** : Barro et Sala-i-Martin.

n	x	ϕ	σ	α	ε	δ	ρ
0,01	0,02	0,02	3/2	1/3	6/5	0,05	0,02

TAB. 1.2 – **Valeurs des paramètres pour la calibration.**

r^*	μ	γ	$\frac{\alpha\varepsilon}{1-\alpha}\gamma^{-1}$	ε^*	$\frac{\hat{c}^*}{\hat{k}^*}$
0,050	0,0808	-0,64	-0,937	1,272	0,220

TAB. 1.3 – **Implications de la calibration.**

5 Annexes de la section 3

5.1 Dérivation de la dynamique optimale et calibration

Posons la fonction valeur,

$$\mathcal{V}(K_t, \mathcal{H}_t) = \underset{\{C_{t+\tau}, \tau \in \mathbb{N}\}}{\mathcal{M}ax} \quad \mathbb{E}_t \left[\sum_{\tau=0}^{\infty} \beta^{*\tau} u(C_{t+\tau}) \right]$$
$$\underline{sc}\ K_{t+1} = (1-\delta)K_t + F(K_t, A_t L_t) - C_t$$

l'utilité du ménage représentatif à l'optimum. $\mathcal{H}_t$ représente l'ensemble d'information à la date t. De façon équivalente, on a,

$$\mathcal{V}(K_t, \mathcal{H}_t) = \underset{\{C_t\}}{\mathcal{M}ax}\ u(C_t) + \beta^* \mathbb{E}_t \left[\mathcal{V}\left((1-\delta)K_t + F(K_t, A_t L_t) - C_t, \mathcal{H}_{t+1} \right) \right]$$

La condition d'optimalité est donnée par,

$$u'(C_t) = \beta^* \mathbb{E}_t \left[\mathcal{V}_1(K_{t+1}, \mathcal{H}_{t+1}) \right]$$

La condition de l'enveloppe,

$$\mathcal{V}_1(K_t, \mathcal{H}_t) = \beta^* \mathbb{E}_t \left[\mathcal{V}_1(K_{t+1}, \mathcal{H}_{t+1}) \left\{ F_1(K_t, A_t L_t) + 1 - \delta \right\} \right]$$

On pose $\lambda_t \equiv \mathbb{E}\left[V_1(K_{t+1}, \mathcal{H}_{t+1}) \right]$ la valeur marginale anticipée d'une unité supplémentaire de capital physique à la période suivante. La condition de transversalité s'écrit alors de la façon suivante,

$$\lim_{s \to \infty} \beta^{*s} \mathbb{E}_t \left[\lambda_s K_{s+1} \right] = 0$$

On note $\Lambda_t = \beta^* \lambda_t$ la valeur marginale anticipée actualisée d'une unité supplémentaire de capital physique à la période suivante.
Finalement la dynamique optimale de cette économie est caractérisée par le

système suivant,

$$\begin{cases} u'(C_t) & = \Lambda_t \\ \Lambda_t & = \beta^* \mathbb{E}_t \left[\Lambda_{t+1} \left\{ F_1(K_{t+1}, A_{t+1}L_{t+1}) + 1 - \delta \right\} \right] \\ K_{t+1} & = F(K_t, A_t L_t) + (1-\delta)K_t \\ \lim\limits_{s\to\infty} \beta^{*s} \mathbb{E}_t \left[\lambda_s K_{s+1} \right] & = 0 \end{cases}$$

On obtient l'équation d'Euler en considérant les deux premières équations du système.

L'état stationnaire déterministe du système dynamique est défini par les égalités suivantes,

$$\hat{y}^* = F(\hat{k}^*, 1) \tag{1.46}$$

$$(n+g+\delta+ng)\hat{k}^* = F(\hat{k}^*, 1) - \hat{c}^* \tag{1.47}$$

$$1 + r^* = \beta^{-1} \text{ par l'équation d'Euler.} \tag{1.48}$$

$$\frac{\hat{y}^*}{\hat{k}^*} = \frac{r^* + \delta}{\alpha} \tag{1.49}$$

$$\frac{\hat{c}^*}{\hat{y}^*} = \frac{\hat{c}^*}{\hat{k}^*} \left(\frac{\hat{y}^*}{\hat{k}^*} \right)^{-1} = 1 - \frac{\alpha}{r^* + \delta}(n+g+\delta+ng) \tag{1.50}$$

$$\frac{\hat{i}^*}{\hat{y}^*} = \frac{\alpha}{r^* + \delta}(n+g+\delta+ng) \tag{1.51}$$

On a calibré ce modèle sur des données américaines. Le taux d'intérêt à l'état stationnaire et les paramètres n,δ sont obtenus en considérant respectivement les données mensuelles de la fred© (on a utilisé la série dénommée *One Year Tresuary Bill Rate (second market)* et obtenu un taux d'intérêt réel en utilisant l'indice des prix à la consommation (*CPI for all urban consummer*) de 1959 à 2001), les données mensuelles de la population totale (récupérée sur le même site) et les données annuelles des *Fixed Assets Tables* du BEA[44].

[44]Le BEA fournit des données agrégées sur le stock de capital, la dépréciation et l'investissement (en nominal, réel et coûts historiques). Pour construire le stock de capital agrégé ils ont utilisé une méthode d'inventaire perpétuel sur des données désagrégées. Pour chaque type d'investissement ils ont considéré un taux de dépréciation spécifique constant dans la chronologie (se reporter à Fraumeni [59] pour une description détaillée). Ainsi le taux de dépréciation (implicite) agrégé n'est pas constant dans la chronologie, il dépend de la composition de l'investissement total (à partir de la fin des années 60 on observe

Pour calibrer le modèle, on a considéré $r^* = 6,13\%$, $n = 1,2\%$ et $\delta = 0.031$. En utilisant l'équation d'Euler à l'état stationnaire on trouve que le facteur d'escompte psychologique est $\beta = 0,94$. Pour calibrer la part du capital physique α et le taux de croissance des variables par tête à l'état stationnaire (g) on a utilisé les restrictions (1.49) et (1.51). Les ratios $\hat{i}^*/\hat{y}^*$ et $\hat{y}^*/\hat{k}^*$ sont évalués à partir des *National Income and Product Accounts Tables* du BEA. On obtient $\alpha = 0,3$ et $g = 0,6\%$. On a posé $\rho = 0,98$, l'inverse de l'élasticité de substitution intertemporelle $\sigma = 3/2$ et l'élasticité de substitution entre les facteurs $\varepsilon = 1,2$. On ne dispose pas d'évidence empirique très claires sur les deux derniers paramètres. Les études empiriques trouvent généralement une élasticité de substitution intertemporelle entre zéro et un (plutôt proche de 0) et une élasticité de substitution entre capital et travail supérieure à l'unité. Pour ces valeurs des paramètres fondamentaux, on obtient les élasticités partielles suivantes,

η_{cx}	η_{ck}	η_{kx}	η_{kk}
3,593	0,436	-0,95	0,927

soit une vitesse de convergence néo-classique de 7,3% contre une vitesse de rattrapage technologique de 2%. Pour cette calibration, les coefficients du processus ARMA(2,2) décrivant le taux de croissance du produit par travailleur sont donnés par,

AR1	AR2	MA1	MA2
1,9073	-0,9088	-1,2847	0,3869

Les racines du polynôme caractéristique associé à la partie MA sont 0,482

une augmentation constante du taux de dépréciation). On a donc considéré le taux de dépréciation implicite moyen sur la période 1929-2000. Il faut cependant souligner que cette façon de procéder n'est pas cohérente avec le modèle. En effet, si sur les données désagrégées la dépréciation est géométrique, cette propriété ne s'étend pas aux données agrégées. On a le même problème avec les Penn World tables de Summers et Heston [146], qui utilisent trois taux de dépréciation selon le type d'investissement les séries de capital obtenues. Dowrick et Rogers [42] utilisent les séries de capital agrégées pour évaluer le modèle de Solow [140] en affirmant que ces données sont cohérentes avec le modèle ; si elles sont effectivement obtenues sur la base d'exercices d'inventaire perpétuels à l'arrivée les séries agrégées (utilisées par les deux auteurs) ne vérifient pas la propriété de dépréciation géométrique et sont donc incohérentes avec le modèle de Solow. Pour que les données soient cohérentes avec le modèle et l'approche empirique il faudrait utiliser les données désagrégées et considérer un modèle de croissance avec différents types de capitaux.

et 0,802. La deuxième racine est relativement proche de l'unité, ceci peut expliquer nos estimations décevantes pour la partie moyenne mobile.

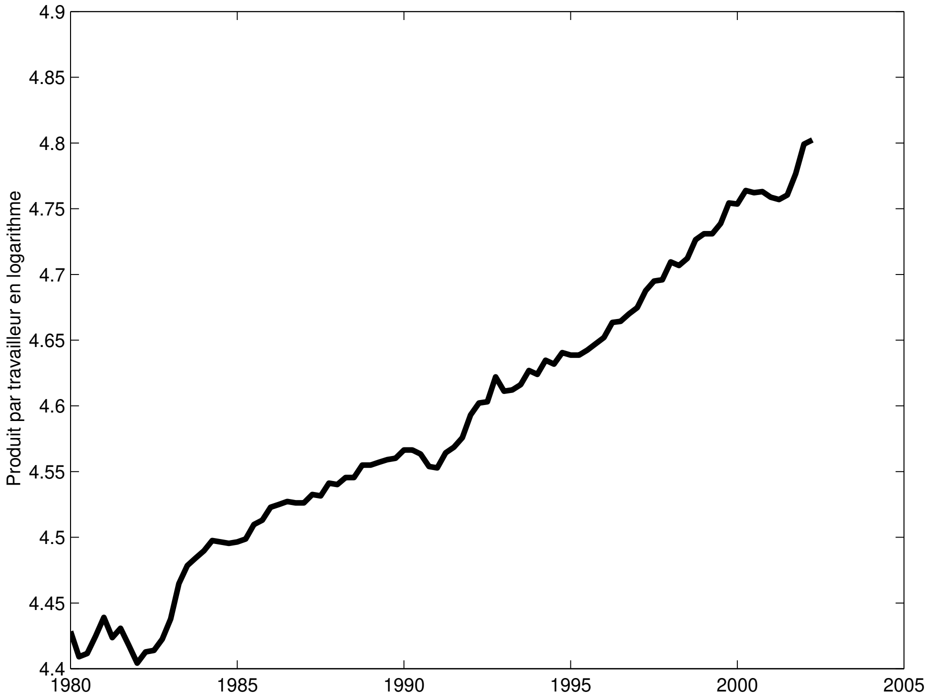

Fig. 1.14 – **Évolution du produit par travailleur en logarithme entre 1980 :01 et 2002 :02 aux États Unis.** Base 100 en 1992 :01.

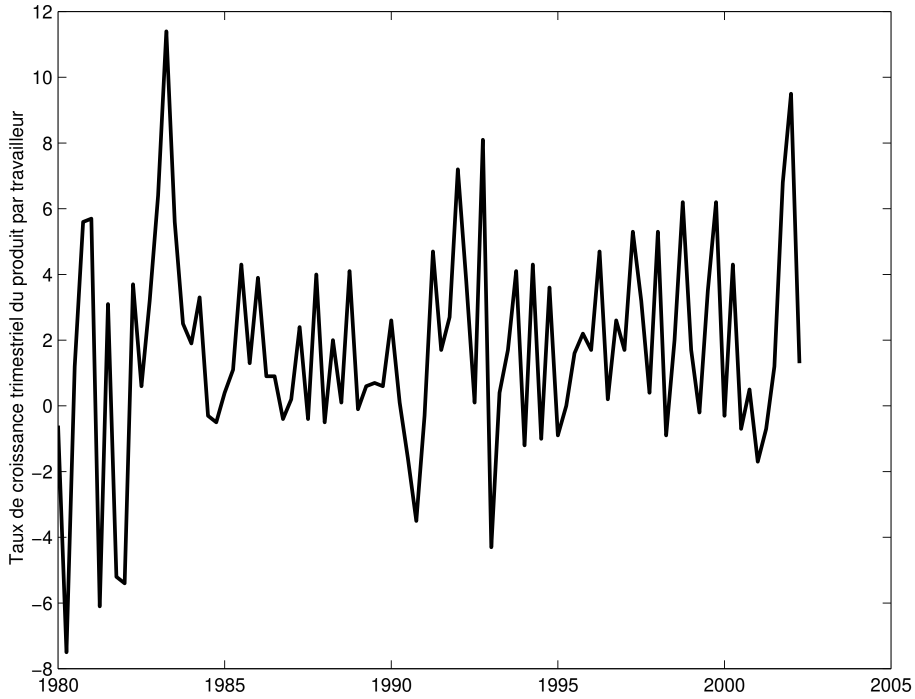

Fig. 1.15 – **Taux de croissance trimestriel annualisé du produit par travailleur.**

Chapitre 2

Convergence déterministe

1 Introduction

L'APPROCHE PAR LES SÉRIES TEMPORELLES conclut rarement à la convergence[1]. A l'exception notable d'Evans et Karras [52], l'hypothèse de convergence n'est jamais admise sur de larges coupes d'économies. Ces résultats peuvent s'expliquer par l'incapacité de cette approche à rendre compte d'une longue phase de transition des économies. Cette limite de l'approche par les séries temporelles apparaît déjà dans Bernard et Durlauf [20]. Les deux auteurs affirment que celle-ci n'est valide que si les économies évoluent le long de leurs états stationnaires respectifs. Un processus stationnaire ne saurait, en effet, rendre compte d'une dynamique de transition importante.

Dans un modèle de croissance néo-classique, du fait de la décroissance du rendement du capital, chaque économie est attirée par une tendance dont la pente est déterminée par le progrès technique exogène. Cette impulsion technologique commune à toutes les économies, constitue une tendance dirigeant la dynamique mondiale. Par construction, une économie — imaginaire — moyenne capture cette tendance. En considérant le produit par travailleur en déviation au produit par travailleur moyen dans la coupe, on élimine la tendance commune et on obtient alors un processus stationnaire. Dans la suite on adopte cette transformation des données pour tester l'hypothèse de convergence.

[1] Ce chapitre est essentiellement basé sur Adjemian [5].

Définition 1.1. *On accepte l'hypothèse de convergence des économies, contre la divergence, si les processus de productivité en déviation à la moyenne en coupe sont stationnaires.*

Cette définition de la convergence peut donc correspondre à la prédiction du modèle néo-classique, elle est vérifiée lorsque les économies évoluent le long de sentiers de croissance parallèles et lorsque les chocs affectant les écarts entre les économies ne persistent pas indéfiniment. En suivant Carlino et Mills [28], on distingue une convergence conditionnelle et une convergence absolue. Cette dernière apparaît lorsque les revenus par tête en déviation à la moyenne spatiale sont stationnaires de moyennes nulles, c'est à dire lorsque les économies ne peuvent s'éloigner les unes des autres durablement. La convergence conditionnelle correspond à une stationnarité en niveau des productivités en déviation à la moyenne spatiale, c'est à dire lorsque les écarts entre les économies ne peuvent augmenter durablement[2].

Tester la convergence revient alors à tester la stationnarité stochastique (l'absence de racine unitaire) *et* la stationnarité déterministe (l'égalité des taux de croissance de long terme des productivités) des écarts de productivité.

Lorsque l'écart initial du produit par tête à sa cible de long terme est trop important, au sens où les données ne sont pas cohérentes avec une linéarisation de la dynamique d'un modèle de croissance autour de l'état stationnaire, le retour à l'état stationnaire n'est pas géométrique et une procédure de test tend à assimiler cette distance à un choc permanent. Ainsi les tests tendent à conclure en défaveur de la convergence.
La figure [2.1] montre l'évolution de la variance en coupe du produit par travailleur pour dix économies européennes observées entre 1950 et 1990. On observe une réduction importante de la variance en coupe ; entre 1950 et la fin des années 70 cette variance est réduite d'un facteur 4. Clairement ces économies vérifient la définition de la σ-convergence. Rappelons que cette

[2]Cette notion de convergence conditionnelle ne correspond pas, a priori, à la définition donnée par Sala-i-Martin [127], puisque rien n'assure que l'hétérogénéité des niveaux de long terme soit liée aux fondamentaux respectifs des économies.

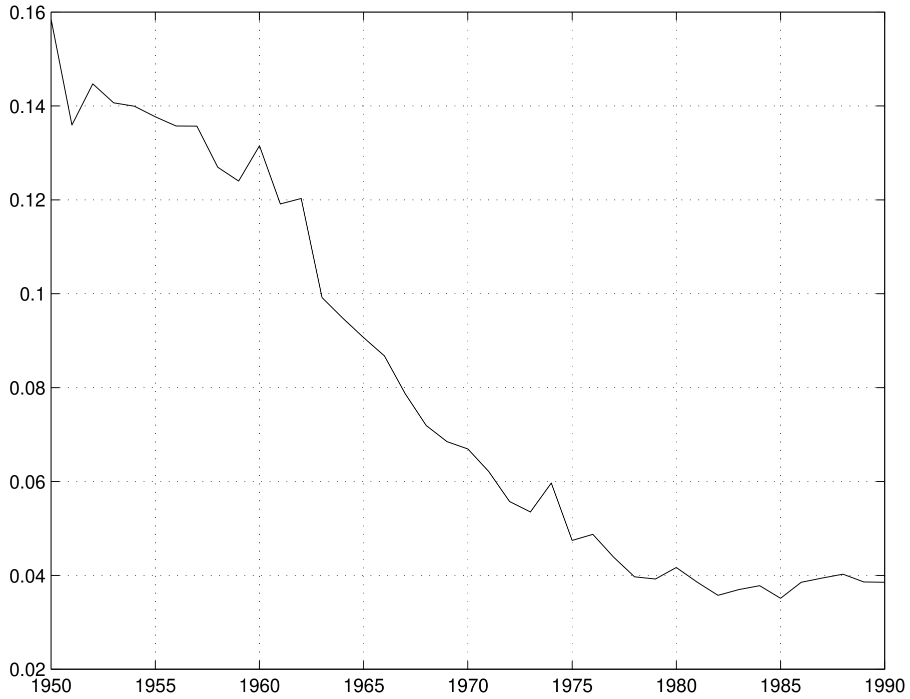

Fig. 2.1 – **Évolution de la variance en coupe de la productivité relative pour dix économies européennes de 1950 à 1990.**

définition induit l'existence d'une force de rappel résultant dans un rapprochement des nations européennes. Malgré la remarquable réduction de la variance en coupe, l'approche par les séries temporelles nous amènerait à conclure à la divergence des économies européennes. Ainsi, un rattrapage important des économies résulte en un rejet de la convergence au sens de la définition [1.1]. Cette conclusion contre intuitive ne s'explique que par l'incapacité d'un processus stationnaire en niveau à reproduire une longue phase de transition.

Ce chapitre montre qu'il est possible *de réconcilier une approche par les séries temporelles et l'existence d'une longue phase de transition des économies*. L'idée est simplement d'associer une part de la dynamique de convergence à la partie déterministe du processus.

La problématique envisagée ici rejoint le thème de la réconciliation des approches par les propriétés des séries temporelles et les tests de β – convergence. Carlino et Mills [28], qui étudient la convergence des régions américaines, proposent d'introduire une rupture dans la partie déterministe des écarts entre les régions, afin de montrer que les chocs sur les écarts ne persistent pas indéfiniment (*Cf.* la problématique de Perron [108]). Ainsi la partie déterministe segmentée du processus absorbe le phénomène de transition, qui pourrait autrement amener à conclure à l'existence d'un choc permanent et donc à la divergence des économies. Ce chapitre propose un pas supplémentaire dans cette voie, en testant l'existence d'une date à partir de laquelle les économies évoluent le long de sentiers parallèles ou le long d'un sentier de croissance commun. La dynamique des économies se décompose alors en deux périodes, une première période de transition, se manifestant par une convergence ou une divergence (déterministe) des économies, puis une période finale où les économies évoluent sur des sentiers de croissance parallèles (ou confondus) et vérifient la définition de la convergence en termes de propriétés des séries temporelles. Sur données européennes (1950-1990), il apparaît que l'on ne peut rejeter l'hypothèse de convergence absolue de des économies.

La section [2] présente rapidement les résultats défavorables à la convergence régulièrement obtenus dans cette littérature puis se concentre sur le cas d'Evans et Karras [52] en montrant que leur résultat favorable à la convergence est vraisemblablement lié aux faiblesses de leur stratégie de test. La section [3] présente une version multivariée du test de stationnarité de Kwiatkowski, Phillips, Schmidt et Shin [86]. On utilise ce test pour étudier la convergence du produit par travailleur parmi dix économies européennes. Il apparaît que le produit par travailleur relatif est non stationnaire au sens déterministe. La section [4] propose un amendement à la définition de la convergence et une stratégie empirique pour tester cette nouvelle définition. On applique cette stratégie sur la même collection d'économies européennes. La section [5] conclut et la section [6] regroupe les annexes. Les graphiques et tableaux sont regroupés en fin de chapitre.

2 Une petite biographie critique

Les travaux empiriques abordent cette définition de la convergence par différents biais, en choisissant généralement de tester la divergence contre la convergence. On distingue deux “traditions”. La première, avec les travaux de Bernard et Durlauf [19], cherche à identifier des relations de cointégration entre les processus de revenus par tête. En effet, pour une collection de N économies, sous l'hypothèse de convergence, on doit avoir $N-1$ relations de cointégration de la forme $(1,-1)$ entre, par exemple, chaque économie *follower* et l'économie *leader*. Ainsi, il existe une unique tendance commune dirigeant la dynamique mondiale des revenus par tête. Bernard et Durlauf [19] proposent une application à 15 pays de l'OCDE. Les deux auteurs mettent en évidence l'existence de deux ou trois tendances stochastiques communes. L'hypothèse de convergence n'est donc pas vérifiée, les économies divergent de façons “ordonnées”.

La deuxième approche est plus contrainte. On teste directement la non stationnarité des écarts de revenus par tête, ou, de façon équivalente, la non stationnarité des processus de revenus par tête en déviation à la moyenne des revenus par tête dans la coupe des économies. Si on parvient à rejeter la non stationnarité des écarts de revenu par tête, alors on admet l'existence d'une relation de cointégration $(1,-1)$ entre les économies *follower* et l'économie *leader*. Cette approche repose généralement sur des tests ADF univariés ou des extensions en panel (voir Quah [119], Evans et Karras [52], Bernard et Jones [22]). A l'exception d'Evans et Karras [52], qui obtiennent un résultat de convergence conditionnelle des revenus par travailleur pour 54 économies de 1950 à 1990, et de Bernard et Jones [22], qui montrent la convergence des Productivités Totales des Facteurs pour 14 économies de l'OCDE de 1970 à 1987, les résultats sont défavorables à la convergence.

Pour expliquer ces conclusions généralement négatives, relativement aux résultats obtenus par l'*approche classique*, Bernard et Durlauf [20] montrent que cette définition exige que les économies soient à leurs états stationnaires respectifs. Cette hypothèse forte ne semble pas acceptable pour l'ensemble des échantillons considérés dans cette littérature. Dès lors, l'incapacité à rejeter la divergence des économies paraît moins étonnante.

2.1 Limites d'une approche en panel

La limite mise en évidence par Bernard et Durlauf n'est pas l'unique problème soulevé par l'approche en série temporelle. On peut voir que les stratégies de tests reposant sur une écriture de type ADF des écarts de productivité ne sont pas adaptées pour tester l'hypothèse de convergence des économies. Cette inadéquation pourrait expliquer le résultat atypique obtenu par Evans et Karras [52]. La procédure proposée par ces deux auteurs repose sur l'écriture suivante de l'évolution de la productivité de l'économie i en écart à la productivité moyenne dans la coupe des économies,

$$\Delta(y_{i,t} - \bar{y}_{.,t}) = \delta_i + \rho_i(y_{i,t-1} - \bar{y}_{.,t-1}) + \sum_{k=1}^{q} \varphi_{i,k}\Delta(y_{i,t-k} - \bar{y}_{.,t-k}) + u_{i,t} \quad (2.1)$$

Si pour tout i on a $\rho_i = 0$, alors les économies divergent (au moins une économie, parmi les N, ne partage pas la tendance commune de croissance). Si pour tout i on a $\rho_i < 0$, les économies convergent. Dans une deuxième étape, les deux auteurs discriminent, s'il y a lieu, entre une convergence absolue et une convergence conditionnelle, en testant l'égalité des δ_i pour $i = 1, ..., N$.

Les deux auteurs proposent de tester $\rho_i = \rho = 0$ dans la dimension panel, en estimant l'équation (2.1), corrigée de l'hétéroscédasticité, et en supposant que les effets individuels sont différents de zéro (sous la nulle les taux de croissance de long terme sont distincts). Ils montrent que la statistique de student associée à l'estimateur du paramètre ρ est, sous la nulle, asymptotiquement distribuée comme une loi normale centrée réduite. Le passage à la dimension panel préserve et augmente les défauts du test ADF. En particulier la présence de paramètres de nuisance dans la distribution de la statistique de test et le traitement confus des tendances déterministes. Par exemple, on peut montrer que la distribution sous la nulle à distance finie dépend de la variabilité dans la coupe des tendances linéaires[3]. Ce paramètre de nuisance complique la mise en œuvre du test, car la distribution du test ne dépend pas que de la taille de l'échantillon. Ainsi, il est nécessaire de simuler la distribution sous la nulle pour chaque échantillon[4].

[3]Voir Adjemian [3].

[4]Evans et Karras [52] ne se risquent pas à un exercice de simulation pour évaluer l'effet de ce paramètre de nuisance sur la puissance et le niveau empiriques de leur test.

Ce test de racine unitaire discrimine entre une non stationnarité stochastique et déterministe (on note $I(1) + T$) et une stationnarité en niveau ($I(0) + c$) pour tous les processus scalaires. La nulle et l'alternative ne peuvent rendre compte du cas où les processus d'intérêt sont stationnaires au sens stochastique et tendanciels au sens déterministe ($I(0) + T$). Dans le cas univarié (se reporter à Stock [144, pages 2783-2784]), si le processus générateur des données est stationnaire autour d'une tendance linéaire (i.e. si $y_t = a + b.t + c.y_{t-1} + \varepsilon_t$ avec $\varepsilon_t \sim iid(0, \sigma^2)$, et $|c| < 1$), alors on montre qu'asymptotiquement l'estimateur de ρ (obtenu en estimant la version univariée de l'équation (2.1)) converge en probabilité vers zéro, car la somme des carrés des résidus est minimisée en différenciant le processus d'intérêt. Asymptotiquement, la présence d'une tendance linéaire amène le test de Dickey-Fuller à accepter la présence d'une racine unitaire (à tort), car la statistique de test ne diverge pas. Mais à distance finie, la réponse de ce test de non stationnarité est moins claire. Tout dépend de l'importance de la tendance linéaire dans le processus générateur de données (i.e. de l'importance du paramètre b par rapport à la *taille* de l'innovation, σ^2). Un simple exercice de simulations montre qu'à distance finie il existe un seuil $\tilde{b}(\sigma^2)$, pour les valeurs possibles de $|b|$, en deçà duquel le test de Dickey-Fuller conclut généralement à la stationnarité[5] (à tort puisque le processus est non stationnaire au sens déterministe). Notons qu'une mauvaise spécification de la partie déterministe tend toujours à favoriser la racine unitaire, quelle que soit la stratégie de test adoptée. La représentation incohérente de la partie déterministe proposée par le test de Dickey-Fuller complique considérablement la mise en œuvre d'une stratégie de test de la divergence.

Cette technologie, mais surtout l'usage qui en est fait par les empiristes de la croissance, semble donc peu convenable pour tester la convergence des économies, car elle ne peut dire clairement pourquoi on rejette ou accepte la divergence des économies, et car elle peut amener à conclure à tort à la convergence des économies au sens de la définition [1.1]. Nous pourrions envisager d'introduire une tendance linéaire dans l'équation (2.1), mais il est plus simple de privilégier directement des technologies de test ne souffrant

[5]Se reporter à West [152]

pas d'un traitement incohérent de la partie déterministe. Même en augmentant le modèle estimé d'une tendance linéaire la procédure de test resterait aussi complexe à cause des paramètres de nuisance.

La démarche d'Evans et Karras [52] pose deux autres questions. La première est liée à l'asymptotique utilisée pour établir la distribution asymptotique sous l'hypothèse nulle. Ces deux auteurs montrent que la statistique de test tend vers une loi normale centrée réduite lorsque la taille de l'échantillon tend vers l'infini avec $\frac{N}{T} \underset{N,T\to\infty}{\longrightarrow} 0$. Ainsi, la dimension temporelle doit diverger plus vite que la dimension individuelle[6]. On voit donc que ce test n'est pas des plus adapté pour l'échantillon considéré ($N = 54$ et $T = 41$) ou pour des échantillons plus larges dans la coupe[7]. La deuxième est liée à la nature de l'hypothèse alternative qui pose que le paramètre ρ_i dans (2.1) doit être négatif pour tout $i \in \{1, \dots, N\}$. Cette approche en panel teste "*Toutes les séries sont à racine unitaire*" contre "*Toutes les séries sont stationnaires*". Il se pourrait que certaines économies convergent alors que d'autres divergent, dans ce cas une approche en panel *homogène* semble difficile à interpréter. Les deux auteurs envisagent eux même cette possibilité,

> "*One might also wish to allow some of the economies to converge while the others diverge and to determine which converge and which diverge. This statistical problem, however, is difficult and well beyond the scope of this paper.*"

puis ils remarquent que le problème ne s'est pas posé pour eux puisqu'ils concluent à la convergence pour les deux échantillons considérés. Il ne s'agit pas seulement d'un problème statistique lié à l'homogénéité du panel, mais aussi d'un problème lié à la représentation de la convergence employée. En considérant le produit par tête en déviation au produit par tête moyen dans la coupe des économies (censé représenté la tendance commune dirigeant la dynamique mondiale sous l'hypothèse alternative) on contraint théoriquement l'approche à être binaire. Si les économies convergent, les processus de produit par tête en déviation à la moyenne individuelle doivent tous être stationnaires. Si au moins une économie diverge (il existe plus d'une ten-

[6] On retrouve la même condition que dans Levin et Lin [88] et [89].

[7] Notons qu'Evans et Karras n'utilisent pas le résultat asymptotique pour mettre en œuvre le test, il s'agit donc plus d'un problème d'élégance.

dance stochastique et/ou déterministe dans le panel) alors les processus de produit par tête en déviation à la moyenne individuelle doivent tous être non stationnaires. Ainsi, la nature binaire des hypothèses nulle et alternative est cohérente avec la lecture des données proposée par Evans et Karras[8].

L'approche proposée par Evans et Karras [52] est insatisfaisante dans plusieurs perspectives. On ne pourrait conseiller l'usage d'une telle approche pour tester l'hypothèse de convergence des économies ou plus généralement tester la non stationnarité d'un panel. On peut maintenant s'interroger sur le résultat donné par ces deux auteurs relativement à la convergence. En effet la conclusion obtenue par Evans et Karras s'oppose aux conclusions généralement reportées dans cette littérature et n'est pas cohérente avec les faits présentés dans l'introduction générale. Cette originalité pourrait bien s'expliquer par les limites évoquées plus haut. Pour étayer cette idée on on peut suggérer l'usage d'une approche alternative sur le même échantillon. Cette approche doit (i) proposer un traitement cohérent de la partie déterministe (même interprétation des paramètres sous la nulle et l'alternative), (ii) offrir une statistique de test libérée des paramètres de nuisance présents dans la statistique proposée par Evans et Karras et (iii) ne pas reposer sur une asymptotique imposant la condition $\frac{N}{T} \underset{N,T\to\infty}{\longrightarrow} 0$. Pour les points ($i$) et ($ii$) on peut utiliser le test de Schmidt et Phillips [130] plutôt que le test de Dickey–Fuller. Pour le point (iii) on a recours à la stratégie de test en panel proposée par Choi [30].

[8]Malgré ses nombreuses limites, la démarche proposée par ces auteurs a un certain succès. Gaulier, Hurlin et Jean-Pierre [61] notent que "*(...) in our study, the rejection of the null hypothesis does not necessarily imply that all economies converge*". Cette assertion est fausse, étant donnée la nature des processus étudiés, si on rejette la présence d'une racine unitaires dans le panel, alors tous les processus de produit par tête en déviation à la moyenne spatiale doivent être stationnaires. Dès lors que l'on considère des données en déviation à la moyenne individuelle, ce cas intermédiaire ne peut exister sous l'alternative. Par contre, l'acceptation de l'hypothèse nulle peut être causée par la divergence d'une seule économie parmi N.

2.2 Une réponse aux limites d'une approche en panel

Plutôt que d'adopter le modèle (2.1), on considère le modèle suivant,

$$
\begin{aligned}
y_{i,t} - \bar{y}_{.,t} &= \mu_i + \delta_i t + x_{i,t} &(2.2)\\
x_{i,t} &= \rho_i x_{i,t-1} + \varepsilon_{i,t} &(2.3)
\end{aligned}
$$

avec $i = 1, ..., N$ les nations et $t = 1, ...T$. Dans ce modèle, utilisé par Schmidt et Phillips [130], la tendance déterministe et la tendance stochastique sont clairement séparées. Ainsi, quelle que soit la valeur du paramètre ρ_i, l'interprétation des paramètres μ_i et δ_i demeure identique. On suppose que $\varepsilon_{i,t}$ est un processus stationnaire de variance $\sigma^2_{\varepsilon,i}$ à partir duquel on peut établir un principe d'invariance, de plus on admet l'indépendance de $\varepsilon_{i,t}$ et $\varepsilon_{j,t'}$ pour tout $i \neq j$. Choi [30] propose une stratégie en panel pour tester $\rho_i = 1$ pour tout i contre $|\rho_i| < 1$ pour au moins un i (s'il n'y a pas d'asymptotique le long de la dimension individuelle) ou contre $|\rho_i| < 1$ pour une fraction de l'échantillon (s'il y a une asymptotique sur N). Cette stratégie est relativement simple à mettre en œuvre. Elle consiste en une agrégation des *p-values* associées aux tests de racine unitaire pour chaque individu. Elle est générale puisque rien n'oblige à considérer le même test pour tous les individus, en particulier on peut adopter des formulations de la partie déterministe spécifiques à chaque individu. Dans cette section on ne considère que le modèle défini par les équations (2.2)–(2.3), et la statistique de student proposée par Schmidt et Phillips [130]. Il s'agit de la statistique de student associée à l'estimateur des moindres carrés ordinaires du paramètre b dans le modèle[9],

$$\Delta\widehat{S}_{i,t} = a + b\widehat{S}_{i,t-1} + u_{i,t} \tag{2.4}$$

avec $\widehat{S}_{i,t} = y_{i,t} - \widehat{\mu}_i - \widehat{\delta}_i t$ et où les paramètres μ_i et δ_i sont estimés sous l'hypothèse nulle $\rho_i = 1$[10]. Schmidt et Phillips montrent que, sous l'hypothèse nulle, $\widehat{S}_{i,t}$ ne dépend pas des paramètres de la partie déterministe du modèle, de la condition initiale de la marche aléatoire et de la variance du terme d'erreur dans (2.2). Les deux auteurs établissent le résultat asymptotique

[9] Voir l'équation (15A) dans Schmidt-Phillips [130, page 261].

[10] L'estimateur $\widehat{\delta}_i$ est obtenu en régressant la différence première $\Delta y_{i,t}$ sur une constante et on a $\widehat{\mu}_i = y_1 - \widehat{\delta}_i$ (on ne peut identifier la condition initiale de la marche aléatoire).

suivant,

$$\tau_i \underset{T\to\infty}{\Longrightarrow} -\frac{1}{2}\left(\int_0^1 \underline{V}_i^2(r)dr\right)^{\frac{1}{2}} \frac{\sigma_{\varepsilon,i}}{\sigma_i}$$

où $\underline{V}(r)$ est un pont brownien centré et σ_i la racine carrée de la variance de long terme associée au processus $\{\varepsilon_{i,t}\}$. Schmidt et Phillips montrent que cette statistique diverge vers $-\infty$ sous l'hypothèse alternative lorsque T tend vers l'infini. Pour chaque nation on estime[11] la statistique de test $\widehat{\tau}_i^* = \widehat{\tau}_i \frac{\widehat{\sigma}_i}{\widehat{\sigma}_{\varepsilon,i}}$ et on estime la *p-value* associée à la statistique de test[12]. Choi [30] considère trois spécifications de la fonction agrégeant les *p-values*. On ne considère ici que la statistique, notée Z, qui d'après les simulations de Choi offre les meilleures propriétés en termes de puissance et niveau empiriques. Cette statistique est définie par,

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(p_i) \tag{2.5}$$

où p_i est la *p-value* associée à l'individu i et $\Phi^{-1} : [0,1] \to \mathbb{R}$ est la réciproque de la fonction de répartition de la loi normale centré réduite. Lorsque T tend vers l'infini, cette statistique est normalement distribuée et elle diverge vers moins l'infini sous l'hypothèse alternative "*Il existe au moins un i tel que $y_{i,t}$ soit stationnaire au sens stochastique*". En effet, lorsque l'hypothèse alternative est vraie, on sait que si T tend vers l'infini la statistique de Schmidt-Phillips tends vers moins l'infini et donc que la *p-value* tend vers zéro. Ainsi la statistique Z tend vers moins l'infini lorsque T tend vers l'infini. Si on considère en plus une asymptotique sur N, alors la statistique Z vérifie les mêmes résultats asymptotiques si on modifie l'hypothèse alternative en considérant l'événement "*Il existe une fraction d'individus tels que $y_{i,t}$ soient stationnaire au sens stochastique*". Cette modification de l'hypothèse alternative est nécessaire pour que la probabilité de violer l'hypothèse nulle soit mesurable asymptotiquement (sinon rien n'assure que le test soit convergent). Choi [30] utilise une asymptotique séquentielle pour établir ces

[11] On utilise un estimateur non paramétrique de la variance de long terme avec une fenêtre *Quadratic Spectral*, le paramètre de lissage optimale, spécifique à chaque nation, est obtenu à l'aide de la procédure automatique proposée par Newey et West [103].

[12] Pour obtenir cette *p-value* on ne considère pas la distribution asymptotique mais une distribution simulée pour $T = 41$.

résultats asymptotiques. On fait d'abord tendre la dimension temporelle vers l'infini puis éventuellement la dimension individuelle. Phillips et Moon [114][13] nous mettent en garde contre ce type d'asymptotique, mais cela ne pose pas réellement problème dans le cas de la statistique Z car on obtient la même distribution asymptotique pour N fini et pour $(T,N)_{seq} \to \infty$.

Sur la base de l'échantillon[14] utilisé par Evans et Karras, on obtient $\widehat{Z} = 25,057 > -1,64$. L'acceptation de l'hypothèse nulle (racine unitaire dans les 54 séries temporelles de produit par travailleur en déviation à la moyenne individuelle) contre l'hypothèse alternative (il existe au moins une série temporelle stationnaire) est très fort[15] (on accepte l'hypothèse nulle quel que soit le seuil du test). La valeur importante de $\widehat{Z}$ s'explique par le fait que les statistiques estimées $\widehat{\tau}_i^*$ sont souvent très proches de zéro[16] et donc les *p-values* estimées souvent unitaires. Les nations de cet échantillon divergent, on retrouve donc une conclusion plus cohérente avec les tests de convergence régulièrement reportés dans la littérature. Ainsi, la conclusion d'Evans et Karras pourrait bien être le résultat des mauvaises propriétés de leur test.

2.3 Prolongements

Une limite du test en panel employé ici est l'hypothèse d'indépendance dans la coupe des perturbations. Dans la suite du chapitre on considère des stratégies qui corrigent de la dépendance spatiale, mais qui ont le défaut de ne pouvoir s'appliquer sur une aussi large coupe d'économies (la dimension

[13] Ces deux auteurs montrent sous quelles conditions une asymptotique séquentielle est équivalente à une asymptotique jointe où les dimensions individuelle et temporelle tendent simultanément vers l'infini (éventuellement avec une restriction sur les vitesses de divergence respectives). Très brièvement, il faut que la convergence en loi de la statistique individuelle vers une distribution individuelle, lorsque la dimension temporelle s'accroît, soit uniforme.

[14] La variable RGDPW dans les *Penn World Tables* 5.6 de Summers et Heston [146] représente le produit réel par travailleur. Cette variable est observée annuellement entre 1950 et 1990 pour 54 économies.

[15] On a vérifié la robustesse de ce résultat au choix de la fonction d'agrégation des *p-values* et à la méthode utilisée pour estimer la variance de long terme.

[16] Elles sont parfois positives.

temporelle doit être beaucoup plus importante que la dimension individuelle). Une façon simple et intuitive de représenter la corrélation spatiale est de penser les processus étudiés comme la somme pondérée de facteurs communs et de composantes idiosyncrasiques. Par exemple, nous pourrions représenter la dépendance individuelle caractérisant le produit par travailleur en déviation au produit par travailleur moyen en adoptant le modèle suivant,

$$y_{i,t} - \bar{y}_{.,t} = d_{i,t} + \lambda_i F_t + \varepsilon_{i,t}$$

où λ_i est un vecteur $1 \times p$, $\{F_t\}_{t=1}^T$ est un processus stochastique multivarié de dimension $p \times 1$ représentant les facteurs communs et $\varepsilon_{i,t}$ est un processus stochastique spécifique à chaque nation orthogonale à $\{F_t\}_{t=1}^T$ et indépendants dans la coupe. Le terme $d_{i,t}$ représente la partie déterministe spécifique à chaque nation. Il s'agit de la représentation employée par Bai et Ng [10]. Ces deux auteurs montrent comment on peut extraire les facteurs communs des données (il s'agit d'une analyse en composante principale sur les données en différence première), puis tester indépendamment de l'ordre d'intégration de $\{F_t\}_{t=1}^T$ la présence d'une racine unitaire dans les processus idiosyncrasiques (éventuellement à l'aide d'une stratégie en panel) mais aussi tester la présence de racines unitaires dans les facteurs communs. On pourrait reprendre le test précédent en corrigeant d'une éventuelle corrélation spatiale à l'aide de l'approche de Bai et Ng. Une problématique bien plus intéressante[17] consisterait en une application de cette approche directement sur les processus de produit par travailleur. Cela nous permettrait d'envisager un exercice à la Bernard et Durlauf [19] étendu sur une large coupe de nations. On peut estimer le nombre de facteurs communs[18] p, tester la non stationnarité de ces facteurs communs puis enfin tester l'existence de relations de cointégration des processus de produit par travailleur avec les facteurs communs en testant la stationnarité des composantes idiosyncrasiques. L'analyse des vecteurs λ_i fournirait alors une *carte* décrivant la dynamique mondiale de croissance. Ce travail reste à faire.

[17] Car il est peu vraisemblable que la conclusion sur la convergence puisse changer.
[18] Voir Bai et Ng [9].

3 Tester la convergence sous la nulle

3.1 Un test de convergence

Dans cette section on propose une stratégie pour tester la convergence sous l'hypothèse nulle en tenant compte d'une éventuelle dépendance dans les dimensions temporelle *et* spatiale.
Puisque l'hypothèse d'intérêt est la convergence, il semble plus naturel de choisir celle-ci sous la nulle. Dans la littérature on choisit plus souvent de placer la divergence sous l'hypothèse nulle (en employant par exemple le test de Dickey-Fuller). Jusqu'à Hobijn et Franses [74][19], les tests de stationnarité ne sont utilisés que pour vérifier la robustesse des conclusions obtenues à l'aide de tests de non stationnarité (divergence) sous l'hypothèse nulle. On tient compte de la dépendance spatiale caractérisant éventuellement les processus de produits relatifs en considérant un test multivarié de stationnarité (une version SUR du test de Kwiatkowski, Phillips, Schmidt et Shin [86]). La correction de la dépendance spatiale (comme pour la dépendance temporelle) est non paramétrique. On n'impose pas de structure spécifique en construisant une distance économique comme dans Conley [32].

Tester la stationnarité dans un cadre multivarié

La version multivariée du test de KPSS [86], repose sur le processus générateur de données suivant[20],

$$\begin{cases} Y_t = \xi.t + R_t + \varepsilon_t \\ R_t = R_{t-1} + u_t,\ R_0 \text{ donné} \end{cases} \tag{2.6}$$

où Y_t, ε_t, u_t et la marche aléatoire R_t sont des processus vectoriels de dimension $N \times 1$, ξ est un vecteur de dimension $N \times 1$. Le processus $\{u_t\}_{t=1}^{T}$ est un bruit blanc multivarié, avec pour matrice de variance-covariance la matrice,

[19]Cette section est proche de la contribution de ces deux auteurs. Je n'avais pas connaissance de leur travail lors de la première écriture de cette section dans mon mémoire de DEA en septembre 1998. Hobijn et Franses emploient une généralisation équivalente du test de KPSS pour réaliser un exercice de clustering sur 112 économies observées entre 1960 et 1989.

[20]Dans la proposition [3.1] qui suit on envisage un cas plus général avec une tendance polynômiale d'ordre p. Le cas considéré dans l'équation suivante suffit pour tester la convergence.

de dimension $N \times N$, Θ. Le processus multivarié $\{\varepsilon_t\}_{t=1}^{T}$ est stationnaire au second ordre, avec pour matrice de variance-covariance Σ, définie positive. Pour rendre compte de processus pouvant être caractérisés par de la dépendance temporelle, on suppose que la perturbation $\{\varepsilon_t\}_{t=1}^{T}$ vérifie les conditions énoncées par Phillips et Durlauf [112, corollaire 2.2 page 476]. Ces conditions sont liées au degré de dépendance temporelle et permettent d'envisager des cas suffisamment généraux (comme, par exemple, une modélisation VARMA de la perturbation). La matrice de variance covariance de long terme, notée Ω, résume alors les structures de dépendances temporelle et spatiale. La condition initiale de la marche aléatoire joue le rôle de la constante dans le processus générateur de Y_t.

Sous l'hypothèse nulle, les N processus scalaires sont stationnaires au sens stochastique. Dans ce cas, la marche aléatoire multivariée, R_t, doit être dégénérée, au sens où $\Theta = 0$. Le processus générateur des données s'écrit alors de la façon suivante,

$$Y_t = R_0 + \xi.t + \varepsilon_t$$

Sous l'hypothèse alternative, au moins un des processus scalaire est non stationnaire au sens stochastique. Dans ce cas la matrice Θ est non nulle dans au moins une direction.

Soit e_t le résidu estimé par les moindres carrés ordinaires du modèle sous l'hypothèse nulle[21], puis $S_t \equiv \sum_{\tau=1}^{t} \varepsilon_\tau$ la somme partielle vectorielle et $\hat{S}_t$ la somme partielle estimée. La version multivariée de la statistique de test de KPSS est donnée par l'expression suivante,

$$\kappa = \text{trace}\left\{ T^{-2}.\hat{\Omega}(l)^{-1}.\sum_{t=1}^{T} \hat{S}_t \hat{S}_t' \right\}$$

où $\hat{\Omega}(l) \equiv T^{-1}\sum_{t=1}^{T} e_t e_t' + T^{-1}\sum_{s=1}^{l} k(s/l) \sum_{t=s+1}^{T}(e_t e_{t-s}' + e_{t-s}e_t')$ est un estimateur de la matrice de variance-covariance de long terme, où $k(x)$ définit la fenêtre pondérant la somme des matrices d'autocorrélation et où le para-

[21]On obtient ce résidu en régressant chaque processus scalaire sur une constante et une tendance linéaire.

mètre de lissage (l) est une fonction de T.

La proposition suivante généralise les résultats de KPSS [86] pour obtenir le comportement asymptotique de κ sous les hypothèses nulle et alternative, en supposant que la partie déterministe est un polynôme d'ordre p[22].

Proposition 3.1. *Sous l'hypothèse nulle, le comportement asymptotique de la statistique κ est donné par*

$$\kappa \underset{l,T\to\infty}{\Longrightarrow} tr\left\{\int_0^1 V_{p+1}(r)V_{p+1}(r)'dr\right\}$$

Sous l'hypothèse alternative (présence de tendances stochastiques dans q processus scalaires) la distribution limite est donnée par,

$$lT^{-1}\kappa \underset{l,T\to\infty}{\Longrightarrow} tr\left\{\left(\int_0^1 \underline{W}_p(r)\underline{W}_p(r)'dr\right)^{-1}\int_0^1\left(\int_0^r \underline{W}_p(s)ds\right)\left(\int_0^r \underline{W}_p(s)ds\right)'dr\right\}$$

où $V_{p+1}(r)$ est un pont brownien d'ordre $p+1$ et de dimension $N\times 1$, $\underline{W}_p(r)$ est un mouvement brownien "détrendé" d'ordre p de dimension $q\times 1$, l est le paramètre de lissage lié à l'estimateur non paramétrique de la matrice de variance-covariance de long terme, $l\sim O_p(T^\delta)$ avec $\delta\in]0,\frac{1}{2}[$.

Preuve Voir la section [6.1, en page 149]. ■

Si sous l'hypothèse nulle les processus étudiés sont stationnaires en niveau ($R_0\neq 0$ et $\xi=0$) la statistique de test de stationnarité κ[23] converge en loi vers une fonctionnelle de pont brownien $V(r)\equiv W(r)-rW(1)$. Si sous l'hypothèse nulle les processus sont stationnaires autour de tendances linéaires ($R_0\neq 0$ et $\xi\neq 0$) la statistique de test de stationnarité κ[24] converge en loi vers une fonctionnelle de pont brownien du second ordre $V_2(r)\equiv W(r)+(2r-3r^2)W(1)+(6r^2-6r)\int_0^1 W(s)ds$. Choi et Ahn [31] ont

[22] Choi et Ahn [31, théorème 2, point (a,*iv*)] proposent un résultat asymptotique sous la nulle, mais il est erroné. Je remercie un rapporteur anonyme d'avoir attiré, volontairement ou non, mon attention sur ce point.

[23] Cette statistique de test est estimée sous l'hypothèse nulle en régressant chaque série sur une constante.

[24] Cette statistique de test est estimée sous l'hypothèse nulle en régressant le produit relatif sur une constante et une tendance linéaire.

montré que la statistique de test diverge si l'ordre de la tendance polynômiale est sous évalué. Par exemple, si on estime la statistique de test en régressant les séries étudiées sur des constantes mais que les processus sont caractérisés par des tendances linéaires, la statistique de test diverge ($\kappa \sim O_p(T^{1-\delta})$). Notons que l'asymptotique ne porte ici que sur la dimension temporelle. On ne peut donc pas appliquer cette procédure sur une large coupe de nations, il faut que la dimension temporelle soit grande par rapport à la dimension individuelle.

La divergence de la statistique de test lorsque l'ordre de la tendance polynômiale est sous évalué motive le choix de Hobijn et Franses [74] dans l'utilisation de la statistique de test κ_0[25], pour tester la convergence des économies. Dans la mesure de l'erreur de seconde espèce, ce test ne conduit pas à accepter à tort l'hypothèse nulle de convergence au sens de la définition [1.1], contrairement aux approches basées sur des représentations de type ADF qui proposent un traitement incohérent de la partie déterministe. Mais en utilisant cette statistique de test, pour tester la stationnarité en niveau sous l'hypothèse nulle, les deux auteurs ne peuvent identifier les raisons qui les poussent à rejeter la convergence. Le rejet de l'hypothèse nulle peut être causé par une non stationnarité stochastique et/ou une non stationnarité déterministe (présence de tendances linéaires dans le processus générateur de données). La procédure proposée ici permet de discriminer entre ces deux causes de non stationnarité.

Une procédure de test de la convergence

L'approche empruntée par Hobijn et Franses [74] n'est pas entièrement satisfaisante. En effet, (i) elle ne délivre pas toute l'information disponible, (ii) elle mène à rejeter la convergence des économies alors que l'aspect tendanciel au sens déterministe des écarts de productivité peut résulter d'un rapprochement des économies. Leurs résultats, défavorables à la convergence, peut s'expliquer par (ii). Pour pallier la propriété (i), on peut utiliser la statistique de test κ_1. Celle-ci permet d'identifier l'origine de la divergence de la statistique de test κ_0. On peut alors décrire une procédure de test de la

[25]On note κ_0 la statistique testant la stationnarité en niveau et κ_1 la statistique testant la stationnarité autour d'une tendance linéaire.

convergence en trois étapes[26] :

Étape (1) Tester la stationnarité stochastique des processus de productivité en déviation à la productivité moyenne dans la coupe des économies, à l'aide de la statistique de test κ_1. Si on rejette la stationnarité, alors les économies divergent car les chocs sur les écarts de productivité ont un effet permanent.

Étape (2) Si ces processus sont stationnaires au sens stochastique, alors on teste la stationnarité déterministe, c'est à dire la nullité du vecteur ξ dans l'équation (2.6). Si on ne rejette pas la stationnarité déterministe, alors les économies convergent au sens de la définition [1.1]. Dans le cas contraire, un phénomène de rattrapage peut expliquer la non stationnarité déterministe.

Étape (3) Si on accepte la convergence, alors on peut discriminer entre une convergence absolue et une convergence conditionnelle en testant la nullité du vecteur R_0 dans l'équation (2.6).

Il convient de noter que la procédure envisagée ici ne repose pas explicitement, comme si nous avions utilisé la technologie ADF, sur un conditionnement négatif du taux de croissance de la productivité relative à son niveau. Ne pas utiliser explicitement ce conditionnement négatif, n'est pas essentiel puisque la définition de la convergence au sens des séries temporelles (*Cf.* Bernard et Durlauf [20]) n'emploie pas explicitement ce mécanisme. Cependant, cette représentation des données ne nous permet pas de calculer une vitesse de convergence des économies, facilitant ainsi la comparaison avec d'autres études[27].

Enfin, notons que cette procédure, contrairement aux approches en panel habituellement employées dans cette littérature, permet de prendre en compte la structure de dépendance spatiale – via la matrice de variance-covariance – liant les économies.

[26] Celles-ci sont appliquées au même processus stochastique multivarié, $\{Z_t\}_{t=1}^T = \{(y_{i,t} - \bar{y}_{.,t})_{i=1}^N\}_{t=1}^T$.

[27] Notons que l'estimation d'une vitesse de convergence dans ce cadre (*ie*, lorsque le rattrapage est capté par la partie déterministe du processus) n'aurait pas grand sens. En effet, si la dynamique observée des économies n'est pas cohérente avec une approximation locale d'un modèle de croissance, la notion même d'une vitesse de convergence (stable dans la chronologie) est remise en cause.

3.2 Le cas européen

On considère le produit réel par travailleur pour dix pays européens[28] de 1950 à 1990 en données annuelles (la variable RGDPW, dans la base de données de Summers et Heston *PWT* 5.6). Avant d'entamer l'exposé des résultats, il convient de remarquer la singularité du modèle complet. En considérant des variables en déviation à la moyenne de l'échantillon, on introduit une singularité puisque la somme (sur les individus) des productivités en déviation à la moyenne spatiale est nulle à chaque date. Pour tester la stationnarité il faut donc retirer une des économies (les résultats sont indépendants du choix de l'économie exclue).

La première étape consiste à régresser, pour chaque économie, la productivité en déviation à la moyenne spatiale sur une constante et une tendance linéaire. En posant le paramètre de lissage $l = 4$ et en employant la fenêtre *Quadratic Spectral* (QS) pour estimer la matrice de variance-covariance de long terme, on obtient $\hat{\kappa}_1 = 0,7574$, à comparer avec la valeur critique au seuil de 5% : $0,8295$ (se reporter au tableau [2.1, en page 163] pour les distributions à distance finie). On accepte donc la stationnarité stochastique des dix processus de produit par travailleur en déviation à la moyenne en coupe, contre la présence d'une racine unitaire dans un des processus scalaires, au seuil de 5%[29].

Pour tester la stationnarité déterministe, c'est à dire la nullité de ξ, on applique un test de Wald d'égalité des scalaires dans ce vecteur. Puisque les résidus (éventuellement autocorrélés) sont tirés dans une loi inconnue, on utilise la technique du Bootstrap séquentiel[30]. On obtient $\hat{W} = 288,108$

[28] La Belgique, le Danemark, la France, l'Allemagne de l'Ouest, la Grèce, l'Irlande, l'Italie, le Luxembourg, les Pays Bas et la Grande Bretagne.

[29] Pour $l = 2,3$ et la même fenêtre, on accepte la stationnarité stochastique avec un niveau au moins supérieur à 2,5%. On obtient respectivement $\hat{\kappa}_1 = 0,6994$ / $0,7574$ pour les fenêtres de Bartlett et QS, lorsque le paramètre de lissage est déterminé par la méthode de Newey et West [103] sans *prewhitening* des résidus (afin d'assurer asymptotiquement la divergence de la statistique de test sous l'alternative).

[30] La technique du bootstrap consiste à tirer de façon équiprobable dans les résidus estimés (contrairement à la technique de Monte Carlo qui utilise des perturbations issues d'un générateur de nombres au hasard) afin de former des échantillons sous l'hypothèse nulle et de calculer la distribution de la statistique de test à distance finie. Cette technique n'est pas valide si les perturbations sont caractérisées par de la dépendance temporelle.

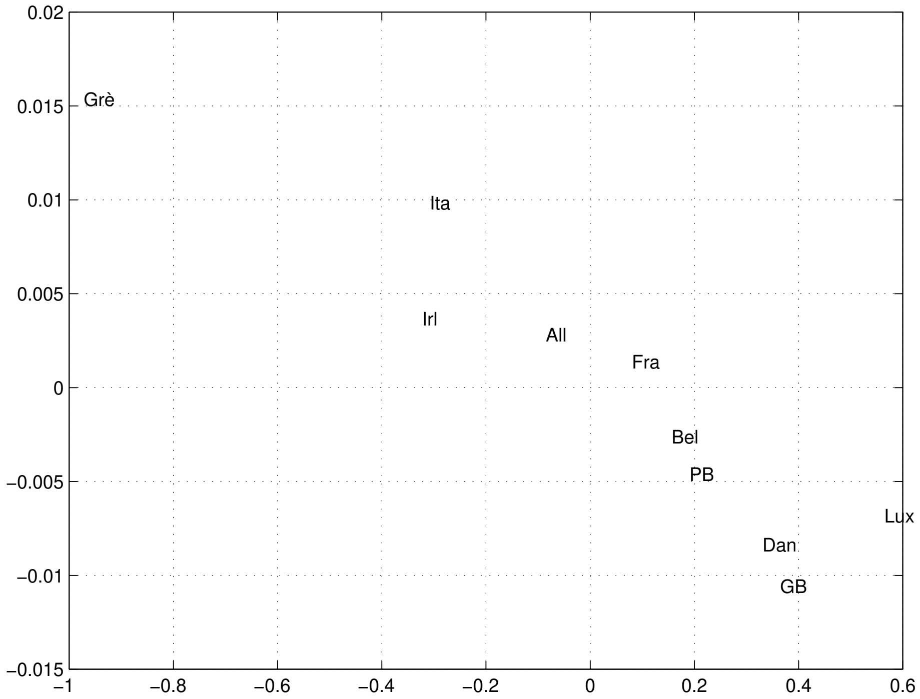

FIG. 2.2 – **Rattrapage des économies européennes.** L'axe des ordonnées donne les coefficients associés à la tendance linéaire dans la régression des écarts de productivité sur un polynôme d'ordre un, l'axe des abscisses donne les conditions initiales en termes de productivité en déviation à la moyenne spatiale.

comme statistique de Wald, et la valeur critique au seuil de 5% est donnée par $39,973$. On rejette donc la stationnarité déterministe contre la non stationnarité déterministe au seuil de 5%.

Il faut s'interroger sur l'origine de ce résultat, afin de conclure sur la divergence ou la convergence des économies. Selon le test de stationnarité stochastique, les chocs sur les écarts entre les économies n'ont qu'un effet transitoire. Selon le test de Wald sur la nullité du vecteur ξ, les économies se

Dans ce dernier cas, on généralise alors cette technique en tirant plusieurs blocs (dont les tailles sont définies de façon aléatoire) de résidus estimés consécutifs pour former chaque échantillon sous l'hypothèse nulle. Il s'agit du bootstrap séquentiel. Pour une description détaillée des généralisations du bootstrap aux processus avec dépendance temporelle, on peut se reporter à Politis, Romano et Wolf [117]. Se reporter à Adjemian [3] pour une présentation précise de la procédure de bootstrap employée ici.

déplacent sur des sentiers de croissance non parallèles. Il semble donc qu'une divergence déterministe (une hétérogénéité des taux de croissance moyens), soit à l'œuvre parmi les dix économies européennes. Cette conclusion, de divergence des économies, est peu satisfaisante ; elle ne résiste pas à la simple observation des graphiques [2.1, en page 121] et [2.2]. Le graphique [2.2] montre clairement que pour chaque économie le coefficient estimé associé à la tendance linéaire ($\hat{\xi}$, représenté sur l'axe des ordonnées) est d'autant plus important que la condition initiale (figurant sur l'axe des abscisses) est faible. Ce constat correspond à l'idée de rattrapage des économies européennes. L'évolution de la variance en coupe (graphique [2.1]) confirme la réduction importante des écarts de productivité en Europe. Ce phénomène ne peut être capturé, de façon satisfaisante par la définition de la convergence et la procédure de test envisagées dans cette section.

C'est l'utilisation de la statistique de test κ_1 plutôt que κ_0 (comme Hobijn et Franses [74]) qui permet cette conclusion nuancée. L'utilisation de la statistique de test de stationnarité stochastique κ_0 amènerait à conclure à la divergence (car les processus d'intérêt sont tendanciels au sens déterministe) mais ne permettrait pas d'associer ce résultat à l'existence d'un éventuel phénomène de rattrapage des économies.

4 Une définition amendée de la convergence

4.1 Une définition amendée

L'objet de cette section est de montrer comment une définition de la convergence basée sur les propriétés des séries temporelles peut être robuste face à l'existence d'un phénomène de rattrapage entre les économies. On peut imaginer le cas où dans une première phase les économies se déplacent le long de sentiers distincts, à des rythmes différents, puis qu'à partir d'une date T_B l'ensemble de ces économies évoluent sur des sentiers de croissance parallèles ou le long d'un sentier commun[31]. Si dans la première phase les économies se rapprochent, alors elle correspond à une phase de rattrapage des économies. Dans ce cas il existe une date à partir de laquelle les économies vérifient la définition de la convergence donnée dans la section [3].

Définition 4.1. *N économies convergent si (i) les écarts de productivité sont stationnaires au sens stochastique, (ii) il existe une date T_B à partir de laquelle ces écarts sont stationnaires au sens déterministe, (iii) antérieurement à cette date on constate un rapprochement des économies. La convergence est absolue si à partir de la date T_B les écarts sont d'espérance nulle.*

On rend compte de l'instabilité des interdépendances entre les économies en ne travaillant que sur la partie déterministe du processus multivarié décrivant l'évolution des productivités en déviation à la moyenne spatiale[32].

[31] En toute généralité la date à partir de laquelle une économie atteint le sentier de croissance commun (ou des sentiers parallèles) n'est pas nécessairement invariante dans la dimension individuelle. L'hypothèse formulée ici n'est pas importante ; une fois que l'on considère la productivité en déviation à la productivité moyenne dans la coupe, la date à partir de laquelle toutes les productivités relatives sont stationnaires est nécessairement commune.

[32] S^t Aubyn [143] propose d'utiliser un filtre de Kalman pour tester la convergence, au sens des séries temporelles, en tenant compte de l'existence d'une phase de transition. Il propose de tester la convergence vers zéro de la variance de l'innovation de la marche aléatoire dans l'équation (2.6) (il travaille dans un cadre univarié et ne prend donc pas en compte la totalité des interdépendances entre les N économies). Cette technologie de test de la convergence lui permet d'envisager l'absence d'invariance dans la partie stochastique du processus, alors que l'on considère ici l'absence d'invariance dans la partie déterministe. Si on pense que le problème de l'approche par les séries temporelles provient d'un écart initial trop grand du produit par travailleur à sa cible de long terme, l'approche de S^t Aubyn est plus difficile à justifier.

Cette méthodologie s'apparente à celle de Carlino et Mills [28], mais ces derniers n'utilisent pas la même technologie de test, ils travaillent dans un cadre univarié et avec des statistiques de tests standards (ADF), qui ne sont pas les plus satisfaisantes pour tester la convergence des économies. De plus, les motivations sont différentes. Ils introduisent des ruptures, en suivant la problématique de Perron [108], pour se donner les moyens de rejeter la présence de racine unitaire dans les écarts de revenus par tête des régions américaines, sans revenir sur la définition de la convergence. Ici, l'introduction d'une rupture commune, en modifiant la définition de la convergence des économies, permet de réconcilier une analyse en séries temporelles avec l'existence d'une longue phase de transition et de montrer que l'aspect tendanciel au sens déterministe des écarts de productivité ne doit pas nécessairement amener à conclure à la divergence des économies.

En notant Z_t le vecteur de dimension $N \times 1$ des productivités relatives (*ie* en déviation à la moyenne spatiale), on adopte la représentation suivante de l'évolution des productivités relatives sous la nulle de stationnarité stochastique,

$$Z_t = \gamma_1 + \gamma_2 \iota(t) + \mu_1 t + \mu_2 \iota(t) t + \varepsilon_t \tag{2.7}$$

avec $\iota(t) \equiv \mathbb{I}_{\{t>T_B\}}(t)$, où $T_B = [\lambda T]$ et $\lambda \in]0,1[$, une fonction indicatrice. Le processus multivarié Z_t se décompose en la somme d'une tendance déterministe segmentée, d'une perturbation stationnaire. L'aspect segmenté de la partie déterministe rend compte de l'instabilité de la croissance. γ_1, γ_2, μ_1, μ_2 sont des vecteurs de dimension $N \times 1$. On distingue trois modèles,

$$\gamma_1 \neq 0, \mu_1 \neq 0, \mu_2 = \gamma_2 = 0 \tag{M0}$$

$$\gamma_1 \neq 0, \gamma_2 \neq 0, \mu_1 \neq 0, \mu_2 = 0 \tag{M1}$$

$$\gamma_1 \neq \gamma_2, \mu_1 \neq \mu_2 \text{ et } \gamma_1, \gamma_2, \mu_1, \mu_2 \neq 0 \tag{M2}$$

Le modèle (M1) envisage le cas où à partir de la date T_B toutes les économies évoluent sur des sentiers de croissance parallèles. Le modèle (M0) envisage le cas où, à un aléa stationnaire près, les économies partagent le même sentier de croissance à partir d'une date T_B. Pour le modèle (M2) l'ensemble des économies n'évoluent pas sur des sentiers de croissance parallèles à partir de cette date.

Pour tester la convergence, absolue ou conditionnelle, tout en intégrant la nature instable de la dynamique de productivité, on peut utiliser la même statistique de test (κ) qu'auparavant. La section suivante a pour objet d'établir ce point.

4.2 Stationnarité et rupture de tendance

Ahn [7] propose plusieurs tests de stationnarité, l'un d'eux correspond à la statistique du test KPSS multivarié, intégrant la possibilité d'existence de ruptures de tendances. Ahn commence par montrer que la statistique du test KPSS diverge lorsque l'on ne prend pas en compte l'existence de telles ruptures, alors même que les processus scalaires sont stationnaires au sens stochastique. Ensuite il donne un théorème décrivant de façon très générale la distribution sous la nulle de la statistique de test. La proposition suivante envisage un cas particulier. Elle donne le comportement asymptotique de la statistique de test lorsque l'on considère les modèles (M0), (M1) ou (M2) sous la nulle. La date de rupture $T_B \equiv [\lambda T]$ est connue.

Proposition 4.2. *(a) On suppose que la perturbation multivariée* $\{\varepsilon_t\}_{t=1}^{T}$ *vérifie les conditions énoncées par Phillips et Durlauf [112]. Sous l'hypothèse nulle (M0), la statistique* κ *adopte le comportement asymptotique suivant,*

$$\kappa_{1,1/0,0}(\lambda) \Longrightarrow trace\left\{\int_0^\lambda \mathcal{H}_2(r,\lambda)\mathcal{H}_2(r,\lambda)'dr + \int_\lambda^1 \mathcal{H}_0(r,\lambda)\mathcal{H}_0(r,\lambda)'dr\right\}$$

Sous l'hypothèse nulle (M1) la statistique κ *adopte le comportement asymptotique suivant,*

$$\kappa_{1,1/1,0}(\lambda) \Longrightarrow trace\left\{\int_0^\lambda \mathcal{H}_2(r,\lambda)\mathcal{H}_2(r,\lambda)'dr + \int_\lambda^1 \mathcal{H}_1(r,\lambda)\mathcal{H}_1(r,\lambda)'dr\right\}$$

Sous l'hypothèse nulle (M2), la statistique κ *adopte le comportement asymp-*

totique suivant,

$$\kappa_{1,1/1,1}(\lambda) \Longrightarrow trace\left\{\int_0^\lambda \mathcal{H}_2(r,\lambda)\mathcal{H}_2(r,\lambda)'dr + \int_\lambda^1 \mathcal{H}_3(r,\lambda)\mathcal{H}_3(r,\lambda)'dr\right\}$$

où les fonctionnelles $\mathcal{H}_i(r,\lambda)$,$i = 0, 1, 2$ et 3, sont définies par :

$$\mathcal{H}_2(r,\lambda) \equiv W(r) - \frac{r}{\lambda}\left(4 - 3\left(\frac{r}{\lambda}\right)^2\right)\int_0^\lambda dW(r) - \frac{6r(r-\lambda)}{\lambda^3}\int_0^\lambda rdW(r)$$

$$\mathcal{H}_0(r,\lambda) \equiv W(r) - \int_0^\lambda dW(r)$$

$$\mathcal{H}_1(r,\lambda) \equiv W(r) - \int_0^1 dW(r) + \frac{1-r}{1-\lambda}\int_\lambda^1 dW(r)$$

et

$$\begin{aligned}\mathcal{H}_3(r,\lambda) \equiv W(r-\lambda) &- \frac{(r-\lambda)\{4(1+\lambda+\lambda^2) - 3(\lambda+1)(r+\lambda)\}}{(1-\lambda)^3}\int_\lambda^1 dB(r)\\ &- \frac{6(r-\lambda)(r-1)}{(1-\lambda)^3}\int_\lambda^1 rdB(r)\end{aligned}$$

(b) Sous l'hypothèse alternative de non stationnarité stochastique, les statistiques de test $\kappa_{./.}(\lambda)$ divergent, $\kappa_{./.}(\lambda) = O_p(T^{1-\delta})$. (c) Si la date de rupture commune λ ne correspond pas à la vraie date de rupture (λ^*), alors $\kappa_{./.}(\lambda) = O_p(T^{1-\delta})$. (d) Si le vrai modèle correspond à (M2) ou (M1), alors $\kappa_{1,1/0,0} = O_p(T^{1-\delta})$. (e) Si le vrai modèle correspond à (M2), alors $\kappa_{1,1/1,0} = O_p(T^{1-\delta})$.

Preuve Voir la section [6.2, en page 156]. ■

On a $\mathcal{H}_3(r,0) \equiv \mathcal{H}_2(r,1) \equiv V_2(r)$, $\mathcal{H}_0(r,0) \equiv W(r)$ et $\mathcal{H}_1(r,0) \equiv V(r)$, on retrouve ainsi les distributions limites données dans la section la proposition 3.1, en l'absence de rupture dans la partie déterministe du processus. La date de rupture, caractérisée par $\lambda \in]0,1[$, intervient comme un paramètre de nuisance. La mise en œuvre du test demeure donc relativement simple. Dès lors que la vraie date de rupture commune est connue, il suffit d'estimer le modèle, (M1) ou (M0), sous l'hypothèse nulle de stationnarité stochastique, puis de calculer la statistique de test κ à partir des résidus estimés[33]. Puisque

[33]Se reporter à la section [6.3, en page 160] pour une discussion sur le niveau et la

λ est un paramètre de nuisance, il faudra retabuler la loi de la statistique de test pour chaque valeur envisagée de la vraie date de rupture.

4.3 Retour sur le cas européen

La statistique de test $\kappa_{1,1/0,0}(\lambda)$, qui correspond au modèle (M0), permet de tester si à partir d'une date T_B, toutes les économies évoluent le long d'un sentier de croissance commun (convergence absolue). La première étape consiste à estimer par les MCO l'équation suivante pour $N-1$ économies,

$$z_{i,t} = y_{i,t} - \bar{y}_{.,t} = \alpha(1-\iota(t)) + \beta(1-\iota(t))t + \varepsilon_{i,t}$$

Pour estimer cette équation, on a considéré la date de rupture $T_B = 1979$ (les résultats de cette estimation sont donnés dans le tableau 2). Ce choix est motivé par l'observation de l'évolution de la variance en coupe des revenus par tête[34] (graphique [2.1, en page 121]). En posant $l = 4$ et en employant la fenêtre QS, on trouve $\hat{\kappa}_{1,1/0,0} = 0.8674$. Au seuil de 5%, on ne parvient pas à rejeter la stationnarité stochastique et l'existence d'une date (1979) à partir de laquelle les productivités évoluent le long d'un sentier de croissance commun contre la non stationnarité stochastique, la convergence conditionnelle, ou la divergence des taux de croissance[35].

Il semble donc que les dynamiques des économies européennes puissent se découper en deux phases. Dans une première phase, les économies sont caractérisées par des dynamiques de transition. Après 1979, elles évoluent le long d'un sentier de croissance commun et vérifient donc à partir de cette date la définition [1.1] de la convergence absolue[36]. La phase de transition

puissance du test à distance finie.

[34]Évidemment, la "constance", à un aléa près, de la variance en coupe après 1979 ne signifie pas que les économies évoluent sur un même sentier de croissance ou sur des sentiers de croissance parallèles.

[35]Pour $l = 2, 3$ on obtient respectivement $\hat{\kappa}_{1,1/0,0} = 1.0469$ et $\hat{\kappa}_{1,1/0,0} = 0.8329$. En utilisant la méthode de Newey et West [103] pour déterminer le paramètre de lissage, on obtient avec les fenêtres QS et de Bartlett respectivement $\hat{\kappa}_{1,1/0,0} = 0.8674$ et $\hat{\kappa}_{1,1/0,0} = 0.8232$.

[36]On a considéré ici une date de rupture exogène. On pourrait endogénéiser le choix de la date de rupture en suivant Zivot et Andrews [155], comme l'indique Ahn [7]. On procède alors en estimant pour chaque date $T^* = [\lambda T]$, avec $\lambda \in [\nu, 1-\nu]$ et $\nu \in]0,1[$, la statistique de test correspondante. On choisit la date de rupture qui maximise la statistique de test

correspond bien à un phénomène de rapprochement des dix économies européennes considérées.

L'introduction de la rupture (modèle (M0)) éclaire les résultats obtenus dans la section ([3.2]). Les tendances linéaires alors obtenues rendent compte de l'instabilité dans la chronologie des dynamiques de croissance respectives des dix économies européennes, elles ne doivent pas nous amener à conclure à la divergence des productivités européennes.

$\kappa_{./.}(\lambda)$, *ie* la date de rupture la plus favorable à l'hypothèse alternative de non stationnarité. Si on note $\mathcal{L}(\lambda)$ la distribution limite de la statistique $\kappa_{./.}(\lambda)$, par le *continuous mapping theorem* on sait que $\sup_{\lambda\in[\nu,1-\nu]} \kappa_{./.,T}(\lambda) \Longrightarrow_{l,T\to\infty} \sup_{\lambda\in[\nu,1-\nu]} \mathcal{L}(\lambda)$. La mise en œuvre de cette stratégie donne des résultats assez peu satisfaisant car la date de rupture est très sensible aux choix sur la procédure d'estimation de la variance de long terme. Voir Adjemian [4] pour une mise en œuvre.

5 Conclusion

Deux conclusions émergent clairement. En premier lieu, l'approche par les séries temporelles conclut à la divergence des économies sur de larges coupes de nations. Le résultat atypique obtenu par Evans et Karras n'est pas, comme ils l'affirment une conséquence de la plus grande puissance de leur test ou du fait qu'ils considèrent la possibilité d'une convergence conditionnelle sous l'alternative[37], mais plutôt le résultat d'une procédure de test assez peu satisfaisante. En second lieu, la définition de la convergence basée sur les propriétés des séries temporelles peut rendre compte de l'existence d'une longue phase de transition, si l'on associe le phénomène de rattrapage à la partie déterministe des processus d'écarts de produit par travailleur. Ainsi, il est possible de conclure à la convergence pour des groupes de nations relativement petits. Pour dix économies européennes, on ne peut rejeter l'hypothèse de l'existence d'une date (1979) après laquelle les économies convergent au sens des séries temporelles, c'est à dire à partir de laquelle les processus de produit par travailleur en déviation à la moyenne spatiale sont stationnaires de moyenne nulle.

Il convient cependant de s'interroger sur la représentation des données employée pour rendre compte de l'existence d'une longue phase de rattrapage des économies. Est-il légitime d'utiliser des tendances linéaires segmentées pour reconnaître ce phénomène alors que le modèle de croissance néoclassique prédit un ajustement "lisse" vers la trajectoire de long terme ? La conclusion pour le cas européen n'est-elle pas influencée par cette approximation ?

Ces questions incitent d'abord à rechercher une formulation plus souple de la partie déterministe, *ie* sans recourir à des tendances linéaires par morceaux, pour rendre compte du processus de rattrapage des économies. Une telle formulation devrait permettre de travailler sur une plus grande variété d'échantillons, puis de reprendre l'exercice de [74] en intégrant l'instabilité du processus de rattrapage des économies. En l'état, avec une représentation sous la forme d'une tendance linéaire segmentée, ce genre d'exercice

[37]Quah [119] ne considère qu'une convergence absolue sous l'alternative

semble difficile à mettre en œuvre à cause du choix de la date de rupture. On pourrait obtenir une trajectoire plus lisse en considérant une tendance polynômiale d'ordre strictement supérieur à l'unité (comme Ouliaris, Park et Phillips [104]) à l'aide de la proposition [3.1, en page 134]. On pourrait aussi, comme Bierens [25], considérer des polynômes de Tchebichev. Le problème est alors de déterminer l'ordre du polynôme, car en se donnant une tendance déterministe suffisamment complexe on parviendra toujours à accepter la stationnarité stochastique[38].

Mais cette recherche d'une spécification plus souple de la partie déterministe ne serait pas réellement beaucoup plus satisfaisante. Filtrer les données à l'aide d'une fonction déterministe plus ou moins complexe, pour éliminer la croissance induite par un rattrapage néo-classique (et/ou technologique) trop important est finalement assez discutable. La démarche idéale serait d'utiliser un modèle qui puisse nous dire ce qui dans la dynamique du produit par travailleur en déviation au produit par travailleur moyen est lié à cet ajustement, puis tester la stationnarité sur la dynamique résiduelle.

[38]Voir Phillips [111] qui montre qu'un processus non stationnaire caractérisé par une racine unitaire peut être représenté de façon équivalente sous la forme d'une fonction déterministe.

6 Annexes

Le lemme suivant présente le principe d'invariance utilisé dans la suite pour dériver les distributions asymptotiques. Il s'agit du corollaire 2.2 dans Phillips et Durlauf [112, page 476].
Avant de présenter ce lemme on rappelle la notion d'α-*mixing* qui permet de caractériser de façon non paramétrique le degré de dépendance d'un processus stochastique. Soient deux tribus F et G. On pose,

$$\alpha(F, G) = \sup_{\mathcal{F} \in F, \mathcal{G} \in G} |\mathbb{P}(\mathcal{F} \cap \mathcal{G}) - \mathbb{P}(\mathcal{F})\mathbb{P}(\mathcal{G})|$$

Si F et G sont indépendantes, alors on doit avoir $\alpha(F, G) = 0$. Posons $\mathcal{F}_a^b = \sigma(\varepsilon_a, \varepsilon_{a+1}, ..., \varepsilon_b)$ la tribu engendrée par $\{\varepsilon_a, \varepsilon_{a+1}, ..., \varepsilon_b\}$, puis

$$\alpha_n = \sup_t \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+n}^\infty)$$

Le processus $\{\varepsilon_t, t \in \mathbb{Z}\}$ est dit fortement mélangeant (*strongly-mixing*) si les coefficients α_n tendent vers zéro lorsque n tend vers l'infini. On caractérise le degré de dépendance d'un processus en spécifiant la vitesse de convergence vers zéro des coefficients α_n.

Lemme 6.1. $\{\varepsilon_t\}_{t=1}^\infty$ *est une suite faiblement stationnaire de variables aléatoires* $N \times 1$ *vérifiant* $\mathbb{E}[\varepsilon_t] = 0$. *Si,*

1. $\mathbb{E}|\varepsilon_t^{(i)}|^\beta < \infty$ *pour une valeur de* β *vérifiant* $2 < \beta < \infty$.
2. $\sum_{n=1}^\infty \alpha_n^{1-\frac{2}{\beta}} < \infty$

alors on peut définir la variance de long terme,

$$\Omega = \lim_{T \to \infty} \mathbb{E}[S_T S_T'] = \mathbb{E}[\varepsilon_1 \varepsilon_1'] + \sum_{k=2}^\infty (\mathbb{E}[\varepsilon_1 \varepsilon_k'] + \mathbb{E}[\varepsilon_1 \varepsilon_k']')$$

Si Ω *est une matrice définie positive, alors*

$$X_T(r) = \frac{1}{\sqrt{T}} \Omega^{-\frac{1}{2}} \sum_{\tau=1}^{[rT]} \varepsilon_\tau \Longrightarrow W(r)$$

un processus de Wiener standard, avec $r \in [0,1]$, α_n est le coefficient α-mixing et où $\Longrightarrow$ indique la convergence en loi lorsque T tend vers l'infini.

La condition 1 correspond au degré d'hétérogénéité autorisé dans le processus multivarié. La condition 2 quant à elle indique le degré de dépendance autorisé. Ce lemme nous dit que l'on peut établir un principe d'invariance si le processus d'intérêt n'est pas trop hétérogène et s'il n'est pas caractérisé par une trop forte dépendance. Ces deux limites sont dépendantes l'une de l'autre. Lorsque β se rapproche de 2, la possibilité d'occurence d'observations aberrantes est plus grande (condition 1) mais les coefficients d'α-mixing doivent retourner plus vite vers zéro (condition 2). Il y a donc un *trade-off* entre hétérogénéité et dépendance.

6.1 Distributions limite du test KPSS multivarié

Distribution limite sous la nulle

Sous l'hypothèse nulle, on a le processus générateur des données suivant,

$$Y_t = R_0 + \xi_1 t + \xi_2 t^2 + \cdots + \xi_p t^p + \varepsilon_t$$

où $\{\varepsilon_t\}_{t=1}^{\infty}$ vérifie les conditions données par Phillips et Durlauf [112] (voir le lemme [6.1]). On peut réécrire de façon plus synthétique le modèle sous l'hypothèse nulle,

$$Y_t = AX_t + \varepsilon_t$$

avec

$$\underset{N\times(p+1)}{A} = \begin{pmatrix} R_0 & \vdots & \xi_1 & \vdots & \ldots & \vdots & \xi_p \end{pmatrix}$$

et

$$X_t = \begin{pmatrix} 1 & t & t^2 & \ldots & t^p \end{pmatrix}'$$

La somme partielle estimée peut alors s'écrire en fonction de la tendance polynômiale et de la perturbation,

$$\hat{S}_{[rT]} = \sum_{t=1}^{[rT]} \varepsilon_t - \sum_{t=1}^{T} \varepsilon_t X_t' \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{[rT]} X_t$$

Posons la matrice diagonale $\delta_T = diag(1, T, ..., T^p)$. On obtient facilement les trois résultats asymptotiques suivants,

$$T^{-1}\delta_T^{-1}\sum_{t=1}^{[rT]} X_t \underset{T\to\infty}{\longrightarrow} \begin{pmatrix} r & \frac{r^2}{2} & \frac{r^3}{3} & \dots & \frac{r^{p+1}}{p+1} \end{pmatrix}' = q(r)$$

$$T^{-1}\delta_T^{-1}\sum_{t=1}^{T} X_t X_t' \delta_T^{-1} \underset{T\to\infty}{\longrightarrow} \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{p} & \frac{1}{p+1} \\ \frac{1}{2} & \frac{1}{3} & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \frac{1}{p+1} & \cdots & \cdots & \cdots & \frac{1}{2p+1} \end{pmatrix} = \int_0^1 g(r)'g(r)dr$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{T} \varepsilon_t X_t' \delta_T^{-1} \Longrightarrow \int_0^1 dW(r)g(r)$$

avec $g(r) = (1, r, ..., r^p)$. Finalement, en intégrant ces résultats asymptotiques dans l'expression analytique de la somme partielle des résidus estimés, on obtient,

$$T^{-\frac{1}{2}}\hat{S}_{[rT]} \Longrightarrow W(r) - \int_0^1 dW(r)g(r)\left(\int_0^1 g(r)'g(r)dr\right)^{-1} q(r) \equiv V_{p+1}(r)$$

Il s'agit d'un pont brownien d'ordre $p+1$ (on reconnaît l'expression donnée par Schmidt et Phillips [130, page 285]). En invoquant le *continuous mapping theorem* et en notant que sous la nulle $\hat{\Omega}(l)$ tend en probabilité vers Ω lorsque $l, T \to \infty$, il vient

$$\kappa \Longrightarrow \mathrm{tr}\left\{\int_0^1 V_{p+1}(r)V_{p+1}(r)'dr\right\}$$

Il est possible de se rapprocher de l'expression donnée par Choi et Ahn [31]. Le terme $\int_0^1 dW(r)g(r)\left(\int_0^1 g(r)'g(r)dr\right)^{-1}$ dans l'expression du pont brownien d'ordre $p+1$, est une matrice $N \times (p+1)$ qui peut être vue, dans l'esprit de Park et Phillips [115], comme l'estimateur de la régression en temps continu

$$dW(r) = \hat{\beta}g(r) + \hat{u}(r) \tag{2.8}$$

avec $\beta = (\beta_0, \beta_1, ..., \beta_p)$. La matrice $\hat{\beta}$ minimise la version en temps continu de la somme des carrés des résidus,

$$\hat{\beta} = \arg\min_{\beta} \int_0^1 \|dW(r) - \beta g(r)\|_2^2 \, dr$$

Choi et Ahn [31] obtiennent une expression similaire dans le lemme B (page 73). La seule différence est que nous obtenons $dW(r)$ là où les deux auteurs notent $W(r)$. Si on suit Choi et Ahn [31] cela revient à dire que la matrice β est obtenue en considérant la régression suivante en temps continu

$$W(r) = \hat{\beta} g(r) + \hat{v}(r) \tag{2.9}$$

qui n'a pas de sens dans le cas du test KPSS (ou SBDH_B selon la notation de [31]) puisque l'on travaille sur des résidus estimés sous l'hypothèse nulle de *stationnarité*. La régression en temps continu doit être "cohérente" avec l'hypothèse nulle.

Distribution limite sous l'alternative

Sous l'hypothèse alternative nous supposons que les $q \leq N$ premiers processus scalaires sont intégrés d'ordre un autour d'une tendance polynomiale d'ordre p.

$$\begin{cases} Y_t^{(i)} = R_t^{(i)} + \xi_1^{(i)} t + \xi_2^{(i)} t^2 + \cdots + \xi_p^{(i)} t^p + \varepsilon_t^{(i)} \text{ pour } i = 1, ..., q \\ Y_t^{(i)} = \xi_1^{(i)} t + \xi_2^{(i)} t^2 + \cdots + \xi_p^{(i)} t^p + \varepsilon_t^{(i)} \text{ pour } i = q+1, ..., N \end{cases}$$

Nous partitionnons le vecteur S_t et la matrice Ω conformément à notre hypothèse alternative,

$$S_t = \begin{pmatrix} S_t^1 \\ S_t^0 \end{pmatrix} \text{ et } \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{10} \\ \Omega_{01} & \Omega_{00} \end{pmatrix}$$

où S_t^1 est un vecteur de dimension $q \times 1$, S_t^0 un vecteur de dimension $(N - q) \times 1$, Ω_{11} est une matrice de dimension $q \times q$, Ω_{00} une matrice de dimension $(N - q) \times (N - q)$ et $\Omega_{01} = \Omega_{10}'$ une matrice de dimension $(N - q) \times q$. De la même façon, on partitionne l'estimateur de la matrice de variance-covariance de long terme, la matrice de variance-covariance de l'innovation de la marche aléatoire multivariée et la somme sur l'échantillon du produit des sommes partielles estimées,

$$\sum_{t=1}^{T} \hat{S}_t \hat{S}_t' = \begin{pmatrix} \sum_{t=1}^{T} \hat{S}_t^1 \hat{S}_t^{1\prime} & \sum_{t=1}^{T} \hat{S}_t^1 \hat{S}_t^{0\prime} \\ \sum_{t=1}^{T} \hat{S}_t^0 \hat{S}_t^{1\prime} & \sum_{t=1}^{T} \hat{S}_t^0 \hat{S}_t^{0\prime} \end{pmatrix} = \begin{pmatrix} \mathcal{A}_{11,T} & \mathcal{A}_{10,T} \\ \mathcal{A}_{01,T} & \mathcal{A}_{00,T} \end{pmatrix} = \mathcal{A}_T$$

Pour montrer la divergence de la statistique de test sous l'hypothèse alternative, nous devons successivement étudier le comportement asymptotique de $\mathcal{A}_T$ puis de $\Omega(l)$.
On a immédiatemment,

$$T^{-2}\mathcal{A}_{00,T} \Longrightarrow \Omega_{00}^{\frac{1}{2}}. \int_0^1 V_{p+1}(r)V_{p+1}(r)'dr.\Omega_{00}^{\frac{1}{2}} \equiv \mathcal{A}_{00}$$

où $V_{p+1}(r)$ est un pont brownien d'ordre $p+1$, de dimension $(N-q)\times 1$, car cela correspond au cas stationnaire. Intéressons-nous à la sous matrice correspondant aux termes intégrés d'ordre un. Ce cas correspond à la "sous régression",

$$Y_t^1 = A^1 X_t + v_t$$

avec $v_t = \varepsilon_t^1 + \sum\limits_{\tau=1}^{t} u_\tau^1$. La somme partielle estimée s'écrit de la façon suivante,

$$\hat{S}_{[rT]} = \sum_{t=1}^{[rT]} v_t - \sum_{t=1}^{T} v_t X_t' \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{[rT]} X_t$$

En notant que $\sum_{t=1}^{T}\sum_{\tau=1}^{t} u_\tau^1 \sim O_p(T^{\frac{3}{2}})$ et $\sum_{t=1}^{T}\sum_{\tau=1}^{t} u_\tau^1 X_t' \delta_T^{-1} \sim O_p(T^{\frac{3}{2}})$, nous avons

$$T^{-\frac{3}{2}}\hat{S}_{[Tr]}^1 \underset{\infty}{\#} T^{-\frac{3}{2}} \sum_{t=1}^{[Tr]}\sum_{\tau=1}^{t} u_\tau^1 - T^{-\frac{3}{2}} \sum_{t=1}^{T}\sum_{\tau=1}^{t} u_t^1 X_t' \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{[rT]} X_t$$

Comme sous la nulle on a,

$$\left(T^{-1}\delta_T^{-1} \sum_{t=1}^{T} X_t X_t' \delta_T^{-1} \right)^{-1} \underset{T\to\infty}{\longrightarrow} \left(\int_0^1 g(s)'g(s)ds \right)^{-1}$$

$$T^{-1}\delta_T^{-1}\sum_{t=1}^{[rT]} X_t \underset{T\to\infty}{\longrightarrow} q(r)$$

où les fonctions $g(\cdot)$ et $q(\cdot)$ sont définies comme dans la section précédente. Puis on a aussi,

$$T^{-\frac{3}{2}}\sum_{t=1}^{[Tr]}\sum_{\tau=1}^{t} u_\tau^1 \Longrightarrow \int_0^r W(s)ds$$

où $W(s)$ est un processus de Wiener de dimension $q \times 1$, et

$$T^{-\frac{3}{2}}\sum_{t=1}^{T}\sum_{\tau=1}^{t} u_\tau^1 X_t' \delta_T^{-1} \Longrightarrow \int_0^1 W(s)g(s)ds$$

Finalement, si on note $T^{-\frac{3}{2}}\hat{S}_{[Tr]}^1 \Longrightarrow \int_0^r \underline{W}_p(s)ds$ et si on remarque que $\int_0^r g(s)'ds = q(r)$, alors on a

$$\underline{W}_p(s) = W(s) - \int_0^1 W(r)g(r)dr\left(\int_0^1 g(r)'g(r)dr\right)^{-1} g(s)'$$

Le terme $\int_0^1 W(r)g(r)dr\left(\int_0^1 g(r)'g(r)dr\right)^{-1}$ peut être obtenu comme l'argument du programme suivant,

$$\min_\beta \int_0^1 \|W(r) - \beta g(r)\|_2^2 \, dr$$

et est donc en rapport avec la régression en temps continu

$$W(r) = \hat{\beta}g(r) + \underline{W}_p(r)$$

C'est pourquoi on parle généralement de mouvement brownien *détrendé* pour la fonctionnelle $\underline{W}_p(r)$. Il s'agit de la projection, au sens de la norme L_2, d'un processus de Wiener sur l'espace engendré par la tendance polynômiale d'ordre p. En regroupant ces résultats asymptotiques, on obtient finalement,

$$T^{-4}\sum_{t=1}^{T}\hat{S}_t^1\hat{S}_t^{1\prime} \underset{\infty}{\#} \sum_{t=1}^{T}\int_{\frac{\tau-1}{t}}^{\frac{\tau}{t}}\left(T^{-\frac{3}{2}\hat{S}}\hat{S}_{[Tr]}^1\right)\left(T^{-\frac{3}{2}\hat{S}}\hat{S}_{[Tr]}^1\right)' dr$$

d'où,

$$T^{-4}\sum_{t=1}^{T}\hat{S}_t^1\hat{S}_t^{1\prime} \underset{T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}}.\int_0^1\left(\int_0^r \underline{W}_p(s)ds\right)\left(\int_0^r \underline{W}_p(s)ds\right)' dr.\Theta_{11}^{\frac{1}{2}} = \mathcal{A}_{11}$$

De façon analogue, on obtient le résultat suivant pour la sous matrice correspondant aux termes produits de processus $I(0)$ et $I(1)$,

$$T^{-3}\sum_{t=1}^{T}\hat{S}_t^1\hat{S}_t^{0\prime} \underset{T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}}.\int_0^1\left(\int_0^r \underline{W}_p(s)ds\right) V_{p+1}(r)'dr.\Omega_{00}^{\frac{1}{2}} = \mathcal{A}_{10}$$

On a donc $\mathcal{A}_{00,T} \sim O_p(T^2)$, $\mathcal{A}_{11,T} \sim O_p(T^4)$ et $\mathcal{A}_{10,T} = \mathcal{A}'_{01,T} \sim O_p(T^3)$, les sous matrices divergent à des rythmes différents puisqu'elles correspondent à des éléments différents (en termes d'ordre d'intégration) du processus multivarié étudié.

On peut maintenant aborder le comportement asymptotique de l'estimateur de la matrice de variance-covariance de long terme. On utilise les résultats de Phillips [110], auquel il faut se reporter pour obtenir plus de détails. On a immédiatement $\Omega_{00}(l) \underset{T\to\infty}{\overset{proba}{\longrightarrow}} \Omega_{00}$ définie positive. L'estimateur de la matrice de variance-covariance de long terme est lié à l'estimateur de la densité spectrale en zéro. En effet,

$$(lT)^{-1}\widehat{\Omega}_{11}(l) = \frac{2\pi}{lT}\hat{f}_{11}(0) = \frac{1}{l}\sum_{s=-l}^{l} k\left(\frac{s}{l}\right)\frac{1}{T}\hat{c}_{11}(s)$$

où $\hat{f}_{11}(0)$ est un estimateur de la densité spectrale en zéro du processus multivarié $v_t^1 = R_t^1 + \varepsilon_t^1$, $k(x) = 1 - |x|$ est la fenêtre de Bartlett (par exemple), et où

$$\hat{c}_{11}(s) = T^{-1}\sum_{t+s=1}^{T} e_t^1 e_{t+s}^{1\prime} = T^{-1}\sum_{t+s=1}^{T}\Delta\hat{S}_t^1\Delta\hat{S}_{t+s}^{1\prime}$$

avec e_t le résidu estimé associé à la régression de Y sur une tendance polynômiale d'ordre p. Étudions le comportement asymptotique de $T^{-1}\hat{c}_{11}(s)$, pour s (positif) donné. On sait que $T^{-\frac{1}{2}} \underset{T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}}\underline{W}_p(r)$, où $\underline{W}_p(r)$ est un processus de Wiener *détrendé* de dimension $p \times 1$. Asymptotiquement, pour s donné, on obtient

$$T^{-1}\hat{c}_{11}(s) \underset{T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}} \int_0^1 \underline{W}_p(r)\underline{W}_p(r)'dr\Theta_{11}^{\frac{1}{2}}$$

En pré-multipliant par la fenêtre de Bartlett puis en sommant sur s on obtient finalement lorsque l et T tendent vers l'infini

$$(lT)^{-1}\widehat{\Omega}_{11}(l) \underset{l,T\to\infty}{\Longrightarrow} \nu\Theta_{11}^{\frac{1}{2}} \int_0^1 \underline{W}_p(r)\underline{W}_p(r)'dr\Theta_{11}^{\frac{1}{2}} = \mathcal{B}_{11}$$

avec $\nu = \int_{-1}^{1} k(s)ds = 1$ pour la fenêtre de Bartlett.
On procède de façon identique pour étudier le comportement asymptotique de $\Omega_{10}(l)$. Soit

$$l^{-1}\widehat{\Omega}_{10}(l) = \frac{2\pi}{l}\hat{f}_{10}(0) = \frac{1}{l}\sum_{s=-l}^{l} k\left(\frac{s}{l}\right)\hat{c}_{10}(s)$$

avec $\hat{c}_{10}(s) = T^{-1}\sum_{t+s=1}^{T} e_{t+s}^{1}e_t^{0\prime}$. Pour s donné, asymptotiquement on obtient

$$\hat{c}_{10}(s) \underset{T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}} \int_0^1 \underline{W}_p(r)dV_{p+1}(r)'\Omega_{00}^{\frac{1}{2}}$$

En pré-multipliant par la fenêtre de Bartlett, puis en sommant sur s, on obtient finalement le résultat suivant

$$l^{-1}\widehat{\Omega}_{10}(l) \underset{l,T\to\infty}{\Longrightarrow} \Theta_{11}^{\frac{1}{2}} \int_0^1 \underline{W}(r)dV_{p+1}(r)'\Omega_{00}^{\frac{1}{2}} = \mathcal{B}_{10}$$

On a donc $\widehat{\Omega}_{00}(l) \sim O_p(1)$, $\widehat{\Omega}_{11}(l) \sim O_p(lT)$ et $\widehat{\Omega}_{10}(l) = \widehat{\Omega}_{01}(l)' \sim O_p(l)$. On dispose des éléments nécessaires pour montrer la divergence de la statistique de test κ sous l'hypothèse alternative et pour donner la distribution limite. On a

$$\kappa = T^{-2}\mathrm{tr}\left\{\widehat{\Omega}(l)^{-1}\mathcal{A}_T\right\}$$

En inversant l'estimateur de la matrice de variance-covariance de long terme partitionné conformément aux hypothèses nulle et alternative[39], on obtient

$$\underset{\infty}{\kappa\#}\mathrm{tr}\left\{\begin{pmatrix} Tl^{-1}\mathcal{B}_{11}^{-1}\mathcal{A}_{11} - l\mathcal{B}_{11}^{-1}\mathcal{B}_{10}\Omega_{00}^{-1} & \bigstar \\ \bigstar & -l\Omega_{00}^{-1}\mathcal{B}_{10}'\mathcal{B}_{11}^{-1} + \Omega_{00}^{-1}\mathcal{A}_{00} \end{pmatrix}\right\}$$

où $\bigstar$ remplace une expression matricielle inutile à notre propos. On peut réécrire cette dernière équivalence asymptotique de façon extensive :

$$\underset{\infty}{\kappa\#}Tl^{-1}\mathrm{tr}\left\{\mathcal{B}_{11}^{-1}\mathcal{A}_{11}\right\} - l\mathrm{tr}\left\{\mathcal{B}_{11}^{-1}\mathcal{B}_{10}\Omega_{00}^{-1}\right\} - l\mathrm{tr}\left\{\Omega_{00}^{-1}\mathcal{B}_{10}'\mathcal{B}_{11}^{-1}\right\} + \mathrm{tr}\left\{\Omega_{00}^{-1}\mathcal{A}_{00}\right\}$$

Finalement on obtient le résultat escompté,

$$lT^{-1}.\kappa \underset{T\to\infty}{\Rightarrow} \mathrm{tr}\left\{\left(\int_0^1 \underline{W}_p(r)\underline{W}_p(r)'dr\right)^{-1}\int_0^1\left(\int_0^r \underline{W}_p(s)ds\right)\left(\int_0^r \underline{W}_p(s)ds\right)'dr\right\}$$

dès lors que $l = O_p(T^\delta)$ avec $\delta \in (0, \frac{1}{2})$.

6.2 Preuve de la proposition 4.2

On ne dérive la distribution limite de la statistique de test que pour le modèle (M1), pour les modèles (M0) et (M2) la preuve est identique. Sous l'hypothèse nulle le processus $\{Y_t\}_{t=1}^T$, s'écrit de la façon suivante :

$$Y_t = Ax_t + \varepsilon_t \tag{2.10}$$

avec $A \equiv (\alpha_1, \alpha_2, \xi)$ une matrice de dimension $N\times 3$, $x_t \equiv (1, \iota(t), (1-\iota(t)t))'$ un vecteur de dimension 3×1. Le processus multivarié $\{\varepsilon_t\}_{t=1}^T$ est stationnaire au second ordre d'espérance nulle, avec pour matrice de variance-covariance Σ. On suppose que ce processus vérifie les conditions énoncées par [112] (voir le lemme [6.1]). On note $\Omega = \Sigma + \sum_{i=1}^{\infty}(\Gamma(i) + \Gamma(i)') = E[\varepsilon_t\varepsilon_t'] + \sum_{i=1}^{\infty}\left\{E[\varepsilon_t\varepsilon_{t+i}'] + E[\varepsilon_{t+i}\varepsilon_t']\right\}$, la matrice de variance covariance de long terme.

La somme partielle estimée, $\hat{S}_t$, par les moindres carrés ordinaires, est donnée par l'expression suivante,

$$\hat{S}_t = \sum_{\tau=1}^{t}\varepsilon_\tau - \sum_{t=1}^{T}\varepsilon_t x_t'\left(\sum_{t=1}^{T}x_t x_t'\right)^{-1}\sum_{\tau=1}^{t}x_\tau$$

[39] Voir [31, Lemme A, page 70] sur l'inversibilité d'une matrice stochastique.

Soit $\delta_T = diag\,\{1, 1, T\}$. On a alors,

$$\delta_T \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} \delta_T = \frac{1}{[\lambda T]([\lambda T]-1)} \times \begin{pmatrix} 2(2[\lambda T]+1) & -2(2[\lambda T]+1) & -6T \\ -2(2[\lambda T]+1) & 2(2[\lambda T]+1) + \frac{[\lambda T]([\lambda T]-1)}{[(1-\lambda)T]} & 6T \\ -6T & 6T & \frac{12T^2}{[\lambda T]+1} \end{pmatrix}$$

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} \varepsilon_t x_t' \delta_T^{-1} \Longrightarrow \begin{pmatrix} \int_0^1 dB(r) & \int_\lambda^1 dB(r) & \int_0^\lambda r dB(r) \end{pmatrix}$$

$$\delta_T^{-1} \sum_{t=1}^{[rT]} x_t = \begin{cases} \begin{pmatrix} [rT] & 0 & \frac{[rT]([rT]+1)}{2T} \end{pmatrix}' & \text{si } r < \lambda \\ \begin{pmatrix} [rT] & [(r-\lambda)T] & \frac{[\lambda T]([\lambda T]+1)}{2T} \end{pmatrix}' & \text{si } r > \lambda \end{cases}$$

La déduction du comportement asymptotique de la somme partielle estimée suit directement,

Cas 1 : $\mathbf{r} < \lambda$

$$T^{-\frac{1}{2}} \hat{S}_{[rT]} \Longrightarrow B(r) - \frac{r}{\lambda}\left(4 - 3\frac{r}{\lambda}\right) \int_0^\lambda dB(r) - \frac{6r(r-\lambda)}{\lambda^3} \int_0^\lambda r dB(r) \quad (2.11)$$

On note $\mathcal{H}_2(r, \lambda)$ cette distribution limite. On a $\mathcal{H}_2(r, 1) = \Omega^{\frac{1}{2}} V_2(r)$, où $V_2(r)$ est un pont brownien du second ordre de dimension $N \times 1$; on retrouve donc le comportement asymptotique de la somme partielle estimée en l'absence de rupture de tendance dans le cas de la stationnarité stochastique autour d'une tendance linéaire.

Cas 2 : $\mathbf{r} > \lambda$

$$T^{-\frac{1}{2}} \hat{S}_{[rT]} \Longrightarrow B(r) - \int_0^1 dB(r) + \frac{1-r}{1-\lambda} \int_\lambda^1 dB(r) \quad (2.12)$$

On note $\mathcal{H}_1(r, \lambda)$ cette distribution limite. On a $\mathcal{H}_1(r, 0) = \Omega^{\frac{1}{2}} V(r)$, où $V(r)$ est un pont brownien de dimension $N \times 1$; on retrouve donc le comportement asymptotique de la somme partielle estimée en l'absence de rupture dans le cas de la stationnarité en niveau.

On s'intéresse à la distribution limite de l'expression suivante,

$$T^{-2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' \underset{\infty}{\#} \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)' dr$$

$$\begin{aligned}\Leftrightarrow T^{-2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' \quad &\underset{\infty}{\#} \quad \int_{0}^{\lambda}\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)' dr \\ &+ \quad \int_{\lambda}^{1}\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)\left(T^{-\frac{1}{2}}\hat{S}_{[rT]}\right)' dr\end{aligned}$$

Par substitution de (2.11) et (2.12) dans l'expression précédente, on obtient,

$$T^{-2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' \Longrightarrow \operatorname{tr}\left\{\Omega\left(\int_{0}^{\lambda}\mathcal{H}_2(r,\lambda)\mathcal{H}_2(r,\lambda)'dr + \int_{\lambda}^{1}\mathcal{H}_1(r,\lambda)\mathcal{H}_1(r,\lambda)'dr\right)\right\}$$

d'où le résultat asymptotique annoncé dans la proposition.

(b) Pour montrer la divergence de la statistique de test sous l'alternative de non stationnarité stochastique, on procède comme pour la preuve de la proposition 4.2. (c) On procède comme pour la preuve du point suivant.

(d) & (e) La statistique de test diverge, dans le cas de stationnarité stochastique, lorsque la partie déterministe du processus multivarié est mal spécifiée sur la deuxième période, c'est à dire lorsque le processus est tendanciel au sens déterministe sur les deux sous-périodes (si on considère $\kappa_{1,1/1,0}(\lambda)$ ou $\kappa_{1,1/0,0}(\lambda)$), ou lorsque le processus est stationnaire en niveau sur la deuxième période (si on considère $\kappa_{1,1/0,0}(\lambda)$). Montrons que la statistique $\kappa_{1,1/1,0}(\lambda)$ diverge lorsque la partie déterministe du processus est mal définie sur la deuxième période (la preuve est identique pour la statistique $\kappa_{1,1/0,0}(\lambda)$). Dans ce cas, on estime modèle (2.10) alors que le processus générateur de données est de la forme suivante,

$$Y_t = \tilde{A}\tilde{x}_t + u_t \tag{2.13}$$

où u_t est un processus stationnaire vérifiant les mêmes conditions que ε_t, avec $\tilde{A} \equiv (\alpha_1, \alpha_2, \xi_1, \xi_2)$ une matrice de dimension $N \times 4$, $\tilde{x}_t \equiv (1, \iota(t), t, \iota(t)t)'$ un vecteur de dimension 4×1. Par substitution de (2.13) dans l'expression analytique du résidu estimé, on obtient,

$$e_t = u_t - \sum_{t=1}^{T} u_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} x_t + \tilde{A}\tilde{x}_t - \sum_{t=1}^{T} \tilde{A}\tilde{x}_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} x_t$$

Soient $\hat{S}_t = \sum_{\tau=1}^{t} e_\tau$, $S_t^u = \sum_{\tau=1}^{t} u_\tau$, $S_t^x = \sum_{\tau=1}^{t} x_\tau$ et $S_t^{\tilde{x}} = \sum_{\tau=1}^{t} \tilde{x}_\tau$. Selon ces notations, on a alors,

$$\begin{aligned} \hat{S}_{[rT]} = S_{[rT]}^u \quad & - \quad \sum_{t=1}^{T} u_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} S_{[rT]}^x \\ & + \quad \tilde{A} \left(S_{[rT]}^{\tilde{x}} - \sum_{t=1}^{T} \tilde{x}_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} S_{[rT]}^x \right) \end{aligned} \qquad (2.14)$$

Asymptotiquement, on a,

- $S_{[rT]}^u = O_p(T^{\frac{1}{2}})$,
- $\sum_{t=1}^{T} u_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} S_{[rT]}^x = O_p(T^{\frac{1}{2}})$,
- $S_{[rT]}^{\tilde{x}} - \sum_{t=1}^{T} \tilde{x}_t x_t' \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1} S_{[rT]}^x = O(T^2)$.

On a donc,

$$\sum_{t=1}^{T} \hat{S}_t \hat{S}_t' = O(T^5)$$

Pour déterminer la vitesse de divergence de la statistique de test, il reste à étudier le comportement asymptotique de l'estimateur de la matrice de variance covariance de long terme. En suivant les notations de Phillips [110],

$$\frac{1}{lT^2}\hat{\Omega}(l) = \frac{2\pi}{lT^2}\hat{f}_{ee}(0) = \frac{1}{l}\sum_{s=-l}^{l} k\left(\frac{s}{l}\right)\frac{1}{T^2}c(s)$$

où $\hat{f}_{ee}(0)$ est l'estimateur de la densité spectrale en zéro du résidu estimé e_t, $k(x)$ est une fenêtre de pondération des matrices d'auto-covariance (la fenêtre de Bartlett par exemple) et où

$$c(s) = T^{-1} \sum_{t+s=1}^{T} e_t e'_{t+s} = T^{-1} \sum_{t+s=1}^{T} \Delta \hat{S}_t \Delta \hat{S}'_{t+s} = O(T^2)$$

On a alors $\hat{\Omega}(l) = O_p(lT^2)$ où $l = O_p(T^\delta)$, avec $\delta \in \left]0, \frac{1}{2}\right[$. Finalement, $\kappa_{1,1/1,0} = O_p\left(T^{1-\delta}\right)$ lorsque le processus sous-jacent est tendanciel au sens déterministe dans les deux sous périodes de l'échantillon.

6.3 Puissance et niveau empiriques du test KPSS avec rupture

Hobijn [73] montrent que la méthode de sélection proposée par Newey et West [103], utilisée sans *prewhitening* des résidus, conduit à une statistique de test convergente. Dans ce cas des simulations montrent que, même pour des petits échantillons, la statistique de test possède un niveau empirique et une puissance empirique satisfaisants. L'introduction de ruptures dans la partie déterministe n'enlève rien aux propriétés à distance finie de ce test.
On considère quatre modèles pour décrire le processus générateur des données sous différentes hypothèses alternatives. Le tableau suivant donne une description de ces modèles,

Modèle 1	$y_t = \xi t + \nu_t$	I(1)+T
Modèle 2	$y_t = \xi t + \varepsilon_t$	I(0)+T
Modèle 3	$y_t = \mu_1 t\iota(t) + \gamma_1 \iota(t) + (1-\iota(t))\gamma_2 + \varepsilon_t$	(M1)
Modèle 4	$y_t = \mu_1 t\iota(t) + \gamma_1 \iota(t) + \varepsilon_t$	(M0)

où ν_t est une marche aléatoire bivariée, ε_t est un processus stationnaire, $\iota(t) = \mathbb{I}_{t \leq T_B}(t)$ avec $T_B = [\lambda T]$ et $\lambda = 0,73$ (comme dans l'application sur les économies Européennes), ξ un vecteur 2×1 de paramètres. On considère que l'innovation de la marche aléatoire est un bruit blanc de variance $\Sigma = I_2$, que le processus stationnaire ε_t est un processus autorégressif d'ordre un, avec pour matrice autorégressive $A = ((a_{i,j}))_i, j = 1,2$ où $a_{1,1} = 0,5$, $a_{1,2} = 0$, $a_{2,1} = 0,2$ et $a_{2,2} = 0,4$, que l'innovation du processus stationnaire est donnée par $\Omega = ((\omega_{i,j}))_{i,j}$ où $\omega_{1,1} = 1$, $\omega_{1,2} = 0,5$, $\omega_{2,1} = 0,5$ et $\omega_{2,2} = 1,5$, que la condition initiale du processus bivarié y_t est nulle.

Les modèles 1,2 et 3 permettent d'évaluer la puissance de la statistique de test $\kappa_{1,1/0,0}$ contre plusieurs alternatives, le modèle 4 permet d'évaluer le niveau empirique de cette même statistique. Le tableau suivant reporte la puissance du test lorsque l'on utilise la procédure de Newey et West [103] et la fenêtre *Quadratic Spectral*[40] pour estimer la matrice de variance covariance de long terme[41].

	Modèle 1	Modèle 2	Modèle 3	Modèle 4
Puissance ou niveau	0,8388	0,6918	0,9659	0,0529

Ce test semble raisonnablement puissant contre les alternatives de non stationnarité stochastique et/ou linéaire, contre la convergence conditionnelle. En tenant compte de la singularité liée à la considération de processus en déviation à la moyenne individuelle, si trois économies sont caractérisées par de la convergence conditionnelle alors on rejette 96 fois sur 100 la convergence absolue.
Cependant, ces résultats en termes de puissance et de niveau sont très sensibles à la spécification du processus générateur des données sous l'alternative. Par l'expression (2.12) on sait que les valeurs des paramètres associés à la tendance et à la constante sous l'alternative constituent des paramètres de nuisance à distance finie (contrairement à la valeur des paramètres sous la nulle). De plus, à distance finie, la distribution de la statistique de test dépend de la *taille de la racine unitaire*, même si asymptotiquement la variance de l'innovation de la marche aléatoire multivariée ne constitue pas un paramètre de nuisance. Dans le cas univarié, Lee [87] montre l'importance de ces paramètres de nuisance à distance finie sur la puissance du test KPSS, lorsque l'on utilise une procédure de sélection du paramètre de troncation basée sur les propriétés de dépendance caractérisant les données. Si ces procédures améliorent le niveau du test, elles détériorent dans de nombreux cas la puissance

[40] Qui selon Andrews [8] et Newey et West [103] donne un estimateur de la variance de long terme plus précis à distance finie.

[41] Les nombres reportés dans ce tableau sont obtenus sur la base de 5000 simulations des modèles 1 à 4 pour un échantillon de taille $T = 50$. Les valeurs critiques sont retabulées par monte-carlo (20000 simulations) pour le cas $N = 2$ et $T = 50$. Le niveau théorique des tests est de 5%.

du test. C'est pourquoi il peut sembler finalement préférable de se donner le paramètre de troncation. Procéder ainsi assure une plus grande puissance du test, mais un biais positif du niveau. C'est l'approche finalement retenue par Hobijn et Franses [74]. Ce choix est aussi légitimé par la nature multivariée de la statistique de test. En effet, sélectionner un paramètre de troncation trop grand par une procédure automatique réduit d'autant plus la puissance que la dimension individuelle du processus multivarié est importante (l'estimation des matrices d'autocorrélation, nécessaires à la construction de la matrice de variance-covariance de long terme est d'autant plus coûteuse en données que la dimension individuelle est importante). Dans le texte on donne les statistiques de test en considérant un paramètre de troncation comme donné ($l = 2, 3, 4$, comme Hobijn et Franses [74]) ce qui induit un niveau empirique trop important, puis confronte ces statistiques avec le cas où le paramètre de troncation est obtenu avec la méthode de Newey et West [103], qui réduit le biais du niveau mais aussi la puissance du test.

Modèle sous la nulle	0,80	0,90	0,95	0,975	0,99
$I(0) + T + c$	0,6974	0,7644	0,8236	0,8785	0,9446
$M(0)$	0,7919	0,8780	0,9518	1,0221	1,1104
$M(1)$	0,4856	0,5259	0,5608	0,5924	0,6292
$M(2)$	0,4143	0,4520	0,4847	0,5136	0,5518

Tab. 2.1 – Les quantiles des distributions pour les différents modèles et statistiques associées sont obtenues par Monte Carlo (20000 simulations), avec $T = 41$, $N = 9$ et $T_B = 1979$. L'issue de l'exercice de monte-carlo ne dépend pas des valeurs des paramètres μ_1, μ_2, γ_1 et γ_2 (voir l'expression analytique de la somme partielle $\widehat{S_t}$).

Pays	$\hat{\alpha}$	$\hat{\beta}$
Belgique	0,21544	$-0{,}0031$
Danemark	0,23437	$-0{,}0091$
France	0,04772	0,0030
Allemagne	$-0{,}01561$	0,0058
Grèce	$-1{,}00579$	0,0200
Irlande	$-0{,}36259$	0,0005
Italie	$-0{,}25835$	0,0114
Luxembourg	0,50287	$-0{,}0110$
Pays Bas	0,30176	$-0{,}0016$
Grande Bretagne	0,34016	$-0{,}0158$

Tab. 2.2 – Estimation des coefficients associés à la constante (α) et à la tendance linéaire (β) par les MCO sur la période (1950-1979).

Chapitre 3

Comptabilité de la croissance

1 Introduction

Dans ce chapitre on s'intéresse plus directement aux contributions respectives de l'accumulation de capital physique et de la technologie à la croissance du produit par tête. Dans le chapitre [1] on a vu comment les régressions à la Barro peuvent nous renseigner sur les sources de la croissance. L'*approche classique* peut, indirectement, nous dire si la dynamique de convergence est plutôt liée au *capital deepening* ou à un retard technologique (*technological catch-up*). On adopte ici une démarche plus traditionnelle pour apporter une réponse à cette question, en estimant des indices de croissance de la Productivité Globale des Facteurs (PGF, par la suite) et en décomposant la croissance du produit par travailleur en la croissance de la PGF et une croissance induite par l'accumulation de facteurs. Cette décomposition doit permettre d'identifier de façon plus fine les contributions des mécanismes néo-classique et technologique de la convergence (ou divergence). En particulier elle nous permettra de montrer que si la PGF est généralement d'une importance toute relative dans la dynamique du produit par travailleur, elle est essentielle pour rendre compte des désastres économiques.

On a vu dans l'introduction générale, que la mesure de la Productivité Globale des Facteurs a une longue histoire. Dans ce chapitre on utilise l'approche *Data Envelopment Analysis* (DEA, par la suite), pour estimer la frontière de production mondiale à chaque date puis dériver des indices de croissance de la PGF. Comme son nom l'indique, cette approche permet d'estimer une frontière de production en construisant une enveloppe convexe dans l'espace des paniers d'*inputs-outputs*. Cette approche a de nombreux avantages, on retient les cinq suivants,

1. Elle fournit un cadre cohérent pour comparer les performances nationales. Toutes les économies sont comparées à une même frontière technologique concave. On peut légitimement voir ce point comme un désavantage, supposer l'existence d'un unique ensemble de production convexe mondiale peut paraître trop fort. On reviendra sur ce point dans la section [4].
2. Elle repose sur un cadre statistique transparent. L'aspect statistique de cette approche est souvent ignoré et n'est devenu évident que depuis le début des années 90. En particulier, on peut montrer sous quelle conditions l'estimateur de la frontière de production est convergent. La faiblesse de cette approche est qu'elle suppose l'absence d'erreur de mesure. Notons néanmoins que les approches envisageant des erreurs de mesures (*stochastic frontier*, voir Kumbhakar et Lovell [85]) supposent que celles-ci portent sur le produit.
3. Elle ne pose aucune hypothèse paramétrique. La technologie n'est pas spécifiée. On suppose seulement que la fonction de production mondiale est croissante en ses arguments et concave. On ne pose pas non plus la loi des distances à la frontière technologique mondiale.
4. Elle ne pose pas d'hypothèse sur la concurrence ou la présence d'externalités. Cet avantage est souvent mis en avant[1], mais n'est pas si évident. L'argument est que l'approche DEA ne nécessitant pas la donnée des prix ou de la part des facteurs dans le produit, ne peut être sensible à ces hypothèses. Cet argument est valable si lorsque l'on parle d'efficience on pense à une efficience technique et non pas à une efficience en toute généralité (qui doit aussi tenir compte de l'optimalité du mélange des *inputs* dans le processus de production, voir Farell [58]).
5. L'indice de croissance de la PGF est décomposable. On peut distinguer deux causes de la croissance de la productivité globale des facteurs : (i) les mouvements de la frontière technologique, *ie* le progrès technique, et

[1]Voir, par exemple, Färe, Grosskopf, Norris et Zhang [56].

(*ii*) les mouvements des unités de production par rapport à la frontière, *ie* la dynamique d'efficience.

Malgré ses indéniables avantages, cette démarche empirique est d'un usage assez peu courant dans les études macro-économiques ; cette démarche est, par exemple, beaucoup plus employée dans l'économie de l'agriculture ou de la banque. Même après la remarquable contribution de Färe, Grosskopf, Norris et Zhang [56] dans l'*American Economic Review*, cette méthodologie est largement ignorée dans le débat sur les sources de la croissance mondiale. Ces quatre auteurs montrent comment obtenir des indices de croissance de la PGF en construisant des frontières de production à l'aide de l'approche DEA. Ils proposent une application à dix-sept économies de l'OCDE entre 1979 et 1988, mais celle-ci demeure purement descriptive[2] et ils n'abordent pas le thème qui nous intéresse. Il faut attendre Kumar et Russell [84] dans la même revue pour que cette approche soit mise au service de la recherche sur les sources de la croissance[3]. Contrairement aux résultats régulièrement reportés par d'autres auteurs (comme Easterly et Levine [48]), ces deux auteurs mettent en avant le rôle essentiel de l'accumulation du capital dans la dynamique du produit par travailleur, pour un échantillon plus large de 57 économies entre 1965 et 1990. La dynamique de la PGF semble plutôt marginale. Les conclusions de Kumar et Russell [84] sont les suivantes : (*i*) il existe un phénomène de rattrapage technologique au sens où les économies ont tendance à se rapprocher de la frontière technologique entre 1965 et 1990 ; (*ii*) le progrès technologique n'est pas neutre, il aurait plus profité aux riches qu'aux pauvres ; et (*iii*) la dynamique du capital peut à elle seule expliquer le caractère *twin peak* de la distribution du produit par travailleur en 1990.

L'approche DEA souffre cependant d'un défaut important : elle est purement statique. La frontière technologique[4] à la date t ne dépend pas de la frontière de production à la date $t-1$. Ainsi on peut se trouver confronté à un problème d'incohérence dynamique dans le choix des technologies, puisque

[2] Ils décrivent plus la méthode que les données.

[3] Le contenu de ce chapitre est assez proche de ce papier et de Henderson et Russell [79] dont je n'avais pas connaissance jusqu'à récemment, je remercie Jérôme Glachant d'avoir porté à mon attention (défaillante) l'existence de ces deux papiers. Les résultats obtenus ici et la présentation sont toutefois sensiblement différents.

[4] La vraie frontière comme la frontière estimée.

rien ne garantit la croissance de la frontière technologique. *La régression technologique n'est pas exclue.* FGNZ [56] estiment un taux de croissance négatif de la technologie pour les États Unis et le Japon entre 1981 et 1982. Kumar et Russell [84] estiment deux ensembles de production non imbriqués pour les années 1990 et 1965 (lorsque la régression technologique est exclue, l'ensemble de production en 1990 doit contenir celui de l'année 1965).

Dans ce chapitre, on propose un amendement du processus générateur des données, implicite dans l'approche DEA, afin d'assurer la croissance de la vraie frontière technologique et une stratégie empirique pour estimer une suite croissante de frontières technologiques concaves. Notre démarche empirique est équivalente à celle proposée par Diewert [38].

La technique utilisée pour exclure l'estimation d'une régression technologique repose sur l'observation qu'une suite d'ensembles de production est croissante si et seulement si les économies sont capables de mélanger les innovations présentes et passées. A la date t l'ensemble de production *dynamiquement cohérent* est la combinaison convexe de son état en $t-1$ et de l'ensemble de production *statique* en t. A chaque date on augmente ainsi l'ensemble de production des nouvelles innovations.
Dans ce chapitre on suppose une propriété de convexité inter-temporelle des choix de technologies, assurant alors la croissance de la suite des ensembles de production. Ainsi, si on peut estimer une suite d'ensembles de production *statiques*, on peut estimer une suite croissante d'ensembles de production.
Si on pense que la régression technologique est impossible, on verra que ne pas imposer la croissance de la suite des ensembles de production résulte dans une sous-estimation de la vraie frontière en 1990, c'est à dire une sur-estimation des distances à la frontière en 1990, et donc une sur-estimation du taux de croissance de l'efficience.

Henderson et Russell [79] estiment, eux aussi, des indices de croissance de la PGF en excluant la régression technologique, mais ils ne considèrent que les dates initiale (1965) et terminale (1990) alors que dans ce chapitre on considère toute l'information disponible entre ces deux années. Contrairement à ces deux auteurs, on formule clairement le processus générateur des données justifiant la démarche empirique, ce qui permet de montrer que

les estimations de la frontière technologique en 1990 reportées par ces deux auteurs sont plus biaisées que les nôtres.
Une autre différence importante est que nous ne considérons pas le capital humain comme un *input* de la fonction de production. Dans notre analyse, l'éducation n'apparaît que dans une seconde étape pour expliquer les distances à la frontière. Notre choix est motivé par la disponibilité des données. Puisque nous proposons une estimation année par année des frontières de production entre 1965 et 1990, il faudrait disposer de séries temporelles du capital humain sur cette période (les données proposées par Barro et Lee [13] couvrent cette période mais ne sont pas annuelles).

On applique cette approche DEA modifiée sur le même échantillon que Kumar et Russell [84]. Comme ces deux auteurs, on ne trouve pas d'évidence forte en faveur de la productivité globale des facteurs. La dynamique d'accumulation du capital suffit pour expliquer le caractère multi-modal de la distribution du produit par travailleur en 1990. Néanmoins la dynamique de la PGF, essentiellement expliquée par l'efficience, *enrichit* le phénomène de divergence. Contrairement à Kumar et Russell [84] on ne trouve pas d'indice en faveur d'un phénomène de rattrapage technologique : en moyenne les économies tendent à s'éloigner de la frontière. Comme se sont essentiellement les économies les plus pauvres qui ont tendance à s'écarter de la frontière, la dynamique d'efficience augmente le phénomène de divergence. De plus il apparaît que les contributions à la croissance ne sont pas invariantes dans la chronologie. Si la dynamique de la PGF peut sembler marginale entre 1965 et 1980, elle apparaît essentielle entre 1980 et 1990. Cette observation est cohérente avec un modèle néo-classique de croissance qui prédit qu'à long terme toute la croissance du produit s'explique par la technologie.

Penser une suite croissante d'ensemble de production, en supposant que les nations peuvent mélanger sans contraintes des ensembles de productions passés, c'est implicitement supposer qu'elles sont déjà capables à une date donnée de mélanger sans contraintes plusieurs technologies, c'est à dire supposer qu'à chaque date l'ensemble de production est convexe. Si on pense la fonction de production mondiale comme un mélange de différentes technologies spécifiques à des groupes de nations, une non convexité de l'ensemble de production mondiale s'interprète comme une impossibilité physique de mélanger certaines technologies. Peut-on mélanger les technologies utilisées par l'Éthiopie avec celles utilisées par les États Unis ? Si la réponse est positive, alors il est légitime de confronter les nations du monde à une unique fonction de production concave. Dans la négative, si l'existence d'une fonction de production mondiale doit avoir un sens, elle ne saurait être concave partout car les technologies sont imparfaitement mélangeables. Ainsi la construction d'une suite d'ensemble de production croissante serait remise en cause. Dans la section [4, en page 207] on teste, sur un échantillon plus large dans la coupe, la convexité de l'ensemble de production en 1995. Cet exercice nous conduit à rejeter l'hypothèse de convexité de l'ensemble de production. Même si cela tend à remettre en cause le travail présenté dans les sections [2] et [3] mais aussi le chapitre [1], ce résultat est intéressant et encourageant dans la mesure où il donne un éclairage original sur le débat convergence locale *versus* convergence globale.

La section 2 présente l'approche DEA en insistant sur le processus générateur des données implicite dans cette approche. On suit la présentation de Banker [11], qui montre la convergence de l'estimateur de la frontière technologique. Dans la même section on amende le processus générateur des données, en supposant que la suite des ensembles de production est croissante, et présente un nouvel estimateur de la frontière. On montre facilement que cet estimateur converge vers la vraie suite de frontière de technologique. La section 3 propose une application sur un large échantillon d'économies entre 1965 et 1990. La section 4 revient sur l'hypothèse de convexité. Les annexes sont reportées en section 5. Les tableaux et graphiques sont situés en fin de chapitre.

2 Estimation non-paramétrique des ensembles de production

Dans cette section on commence par motiver, section [2.1], et présenter, section [2.2], l'approche DEA. La section [2.3] indique comment construire des indices de croissance de la productivité totale des facteurs à partir de l'estimation des frontières technologiques.

2.1 Motivations & Présentation heuristique

On suppose qu'il existe une unique fonction de production à chaque date. Chaque nation $n = 1, \ldots, N$ est confrontée à la même fonction de production concave à une date $t = 0 \ldots, T$. On note $\mathcal{F}_t(\mathbf{x}_{n,t})$ cette fonction, où le vecteur $M \times 1$ $\mathbf{x}_{n,t} \subseteq \mathbb{R}_+^M$ correspond au panier des *inputs* utilisés dans le processus de production (dans le cas étudié ici on a $M = 2$, les *inputs* sont le capital et le travail). Cette fonction de production satisfait les hypothèses habituelles relativement à ses arguments. Le taux de croissance positif de la technologie, c'est à dire la déformation de la fonction de production, est spécifique à chaque nation à travers son choix de panier d'*inputs*. Ainsi on ne pose pas d'hypothèse sur la neutralité du progrès technologique. Puisque l'unicité de la fonction de production est admise, il semble difficile de supposer que toutes les nations soient capables d'exploiter efficacement la frontière technologique définie par la fonction de production. Ainsi, le produit[5] d'une nation n à la date t, $y_{n,t} \in \mathbb{R}_+$, peut être inférieur à $\mathcal{F}_t(\mathbf{x}_{n,t})$. Une façon naturelle de représenter cette incapacité à pleinement exploiter la frontière technologique est d'exprimer le produit de la façon suivante,

$$y_{n,t} = A_{n,t}\mathcal{F}_t(\mathbf{x}_{n,t})$$

avec le niveau d'efficience, $A_{n,t}$, dans l'intervalle $]0, 1]$ pour tout $(n, t) \in \{1, \ldots, N\} \times \{0, \ldots, T\}$. Cette spécification de l'efficience, n'est pas *neutre*. Dans cette formulation, les distances à la frontière technologique sont orthogonales à l'espace engendré par les paniers d'*inputs*. Ainsi l'efficience est neutre au sens de Hicks. L'approche DEA permet, de considérer d'autres

[5] On suppose ici que le produit est un scalaire, mais l'approche DEA s'étend facilement au cas où le processus de production aboutit en plusieurs *outputs*.

directions pour mesurer les distances à la frontière technologique. Ici l'inefficience affecte également les *inputs*, on pourrait par exemple désirer que celle-ci n'affecte qu'un *input* (le travail chez Hall et Jones [69]). On dira qu'une nation n est efficace à la date t si $A_{n,t} = 1$. Il s'agit clairement d'un abus de language car une nation peut se situer sur la frontière en utilisant trop d'*input* ou en produisant trop peu[6].

Si on s'intéresse aux sources de la croissance du produit, on doit séparer la croissance induite par les dynamiques de $A_{n,t}$ et $\mathcal{F}_t(\bullet)$ de celle causée par la dynamique des *inputs*, $\mathbf{x}_{n,t}$. Seuls $y_{n,t}$ et $\mathbf{x}_{n,t}$ sont observés. Afin d'évaluer le rôle de la technologie et de l'efficience dans la croissance, on doit estimer les fonctions de production $\{\mathcal{F}_t(\bullet), t = 0, \dots, T\}$. Les distances à la frontière sont alors obtenues par complémentarité. On peut ainsi construire des indices de croissance de la technologie et de l'efficience, et, en composant ces indices, obtenir un indice de croissance de la productivité globale des facteurs.

Färe, Grosskopf, Norris et Zhang [56] estiment $T+1$ fonctions de production *statiques*. A la date t la frontière technologique est construite comme une enveloppe concave des observations d'*inputs-output* de la date t. On utilise pour cela une enveloppe linéaire par morceaux. Évidemment la frontière n'est pas identifiable en l'état, car on peut construire une infinité de frontières vérifiant la propriété de concavité et enveloppant les observations. Pour rendre ce problème identifiable on sélectionne le plus petit ensemble de production convexe contenant toutes les observations. La figure [3.8, en page 230] représente cette enveloppe linéaire par morceaux dans le cas avec un *input*. On distingue trois cas selon la nature des rendements d'échelle.

Si les rendements sont constants (CRS pour *Constant Returns to Scale*) alors la frontière technologique est représentée par la droite [O,A). Pour tout $\lambda > 0$ si (x, y) est sur la frontière technologique alors $(\lambda x, \lambda y)$ appartient aussi à la frontière. On voit que la frontière est déterminée par une seule observation, le point A, celle qui est caractérisée par la productivité la plus importante c'est à dire le ratio y/x le plus grand. Ceci n'est plus vrai s'il

[6] Pour un exemple, voir la figure 5.F.1 dans Mas-Colell, Whinston et Green [98, page 150].

y a plus d'un *input* dans le processus de production, la frontière est alors déterminée par plus d'une observation.

Si les rendements d'échelle sont non croissants (NIRS pour *Non Increasing Returns to Scale*) la frontière technologique est représentée par la droite segmentée $[O, A] \cup [A, C] \cup [C, D)$. Pour tout $\lambda \in [0, 1]$, si (x, y) se situe sur la frontière technologique alors le couple $(\lambda x, \lambda y)$ appartient à l'ensemble de production. Dans ce cas rien n'assure que le couple transformé demeure sur la frontière. De plus, si λ est supérieur à l'unité rien n'assure que $(\lambda x, \lambda y)$ demeure dans l'ensemble de production. Ainsi les rendements peuvent être constants ou décroissants. Dans l'exemple donné par la figure [3.8, en page 230] on a une zone à rendements constants entre les points O et A, puis une zone à rendements décroissants par la suite[7]. L'approche DEA est une approche locale, *ie* on estime la frontière au dessus de chaque observation, il n'est pas nécessaire de fixer des rendements d'échelle homogènes pour estimer la frontière. Finalement on observe que dans ce cas la frontière est toujours déterminée par plus d'une observation.

FGNZ [56] considèrent enfin le cas où les rendements sont variables (VRS pour *Variable Returns to Scale*), la frontière correspond alors à la droite segmentée $[B, A] \cup [A, C] \cup [C, D)$. Dans ce cas il existe un coût d'entrée dans le processus de production, on ne peut produire une quantité strictement positive si la quantité de facteur est inférieure à B. Cette contrainte introduit une zone à rendements croissants, entre les points B et A, mais elle est de mesure nulle et l'ensemble de production vérifie toujours la propriété de convexité[8]. La différence essentielle avec le cas NIRS est que l'origine n'appartient pas à l'ensemble de production. Ainsi pour tout λ dans l'intervalle $[0, 1]$ si (x, y) est sur la frontière on n'a pas nécessairement $(\lambda x, \lambda y)$ dans l'ensemble de production. Un ensemble de production VRS est défini par une seule contrainte : la convexité.

Finalement, il apparaît que l'on peut classer ces différents ensembles. L'ensemble de production VRS est inclus dans l'ensemble de production

[7] Notons qu'après le point C les rendements sont nuls.

[8] On verra plus loin comment définir un ensemble de production non convexe.

NIRS lui-même inclus dans l'ensemble de production CRS[9].

Une question vient immédiatement à l'esprit : sous quelles conditions cet estimateur, qui pourrait paraître naif, tend vers la vraie frontière technologique lorsque le nombre d'observations tend vers l'infini ? On abordera formellement cette question dans la section suivante. Pour établir la convergence de ce type d'estimateur de la frontière de production, on doit se donner un modèle statistique. Brièvement : il faut en premier lieu définir la frontière et ses propriétés (concavité, croissance, continuité,...), puis en second lieu spécifier comment les nations se répartissent sous la frontière. C'est la partie probabiliste du modèle, on supposera que l'efficience est une variable aléatoire sans formuler explicitement sa densité. En toute généralité, les propriétés asymptotiques de ce genre de modèles sont encore peu connues. Afin d'entrevoir les difficultés et particularités inhérentes à ce genre de modèle on peut rappeler un exemple plus simple, car paramétrique, abordé dans les cours de statistique inductive les plus standards[10].

Soit $(X_1, \ldots, X_N)$ un N-échantillon extrait d'une loi exponentielle décentrée dont la densité est donnée par,

$$f_X(x; \theta) = e^{-(x-\theta)} \mathbb{I}_{\{x \geq \theta\}}$$

Le support de cette loi est donné par $[\theta, \infty[$, on veut estimer la borne θ. La densité jointe du N-échantillon est alors donnée par,

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{i=1}^{N} f_X(x_i; \theta) \\ &= e^{-\sum_{i=1}^{N}(x_i - \theta)} \mathbb{I}_{\{\inf\limits_i x_i \geq \theta\}} \end{aligned}$$

On voit que la vraisemblance est une fonction croissante du paramètre θ. L'estimateur du maximum est défini par,

$$\hat{\theta}_N^{MV} = \inf_i X_i$$

[9] Le lecteur intéressé peut se reporter à Färe Grosskopf et Lovell [55] pour une présentation exhaustive.

[10] On reprend l'exemple donné par Gouriéroux et Monfort [66, pages 183-184].

Clairement, l'estimateur du maximum de vraisemblance est biaisé puisque l'espérance de $\inf_i X_i$ est égale à $1+\theta$. Ce résultat relatif au biais est assez intuitif et il est directement en rapport avec le biais inhérent à l'approche DEA. Puisque toutes les nations se situent sous la frontière technologique, si on estime la frontière concave en considérant l'enveloppe concave minimale des observations on sous-estime nécessairement la vraie frontière technologique. Malgré son biais, cet estimateur est convergent. En effet, on a,

$$\begin{aligned}
\mathbb{P}\left(\left|\hat{\theta}_N^{MV} - \theta\right| > \varepsilon\right) &= \mathbb{P}\left(\hat{\theta}_N^{MV} > \theta + \varepsilon\right) \\
&= \mathbb{P}\left(\inf_i X_i > \theta + \varepsilon\right) \\
&= \mathbb{P}\left(X_i > \theta + \varepsilon ; i \in \{1, \ldots, N\}\right) \\
&= \prod_{i=1}^{N} \mathbb{P}\left(X_i > \theta + \varepsilon\right) \\
&= \left(\int_{\theta+\varepsilon}^{\infty} e^{-(x-\theta)} dx\right)^N \underset{N\to\infty}{\longrightarrow} 0
\end{aligned}$$

où la première égalité s'explique par le biais de l'estimateur du maximum de vraisemblance. Ainsi l'estimateur du maximum de vraisemblance converge en probabilité vers θ. Notons que ce résultat asymptotique ne tient que parce que la loi exponentielle nous assure que pour tout $\eta > 0$ on a $\mathbb{P}(X < \theta + \eta) > 0$. Si nous pouvions trouver une valeur de η telle que $\mathbb{P}(X < \theta + \eta) = 0$, alors l'intégrale sous la puissance N serait unitaire. On ne peut identifier la borne du support de la variable aléatoire X s'il n'est pas possible d'observer des réalisations arbitrairement proches de celle-ci. Ceci nous donne par analogie l'hypothèse clef pour estimer la frontière technologique : il faut que la distribution de l'efficience soit telle que l'on puisse observer des réalisations du produit arbitrairement proches de la frontière.

Une autre "étrangeté" de cet estimateur est qu'il ne converge pas en $\sqrt{N}$. En effet, on a

$$\begin{aligned}
\mathbb{P}\left(\hat{\theta}_N^{MV} - \theta < z\right) &= 1 - \mathbb{P}\left(\hat{\theta}_N^{MV} > z + \theta\right) \\
&= 1 - \mathbb{P}\left(X_1 > z + \theta, X_2 > z + \theta, \ldots, X_N > z + \theta\right) \\
&= 1 - \left(\int_{\theta+z}^{\infty} e^{-(x-\theta)} dx\right)^N \\
&= 1 - e^{-Nz}
\end{aligned}$$

De façon équivalente, on a donc,

$$\mathbb{P}\left(N\left(\hat{\theta}_N^{MV} - \theta\right) < z\right) = 1 - e^{-z}$$

on reconnaît la fonction de répartition de la loi exponentielle. Cette dernière équation nous donne deux informations : (i) l'asymptotique n'est pas normale lorsque N tend vers l'infini et (ii) l'estimateur du maximum de vraisemblance est super-convergent puisqu'on a $\hat{\theta}_N^{MV} - \theta \sim O_p(N^{-1})$ plutôt que $\hat{\theta}_N^{MV} - \theta \sim O_p(N^{-\frac{1}{2}})$.
Pour finir, notons que si $\underset{i}{inf} X_i$ semble être un candidat naturel pour estimer la borne inférieure du support de la variable aléatoire X, il ne correspond pas nécessairement à l'estimateur du maximum de vraisemblance (tout dépend de la loi de X). Pour que cet estimateur soit l'estimateur du maximum de vraisemblance, il faut que la densité de la variable aléatoire X soit monotone décroissante (ce qui est vérifié par la loi exponentielle).

L'estimation d'une frontière technologique et l'étude des propriétés asymptotiques de son estimateur sont bien plus complexes que dans cet exemple. Cette inflation de complexité s'explique essentiellement en notant que (i) l'objet que l'on cherche à estimer n'est pas un simple point mais une fonction non paramétrique[11] de $\mathbb{R}_+^M$ dans $\mathbb{R}_+$ et (ii) contrairement à l'exemple donné plus haut la densité de l'efficience est elle aussi non paramétrique. Simar et Wilson [139] proposent une revue des résultats disponibles aujourd'hui. Outre la convergence de l'estimateur DEA établie par Banker [11], on connaît la vitesse de convergence de cet estimateur. Comme tout estimateur non paramétrique la vitesse de convergence dépend de façon cruciale du nombre

[11] Il s'agit donc d'estimer une infinité de points.

d'*inputs* considérés (*Curse of dimensionality*) et du degré de *smoothness* de la frontière de production. Dans le cas qui nous intéresse, où on supposera la continuité de la fonction de production et où on considérera deux facteurs de production (le travail et le capital), on a une convergence en $\sqrt{N}$ (voir Kneip, Park et Simar [81]). Ce résultat est assez étonnant car on obtient une vitesse de convergence inférieure si on cherche à estimer un modèle non paramétrique standard de la forme $Y_i = m(X_{1,i}, X_{2,i}) + u_i$ avec $\mathbb{E}[u_i] = 0$ (voir Härdle [72] ou Pagan et Ullah [106]). Ce résultat est à mettre en relation avec la super-convergence de l'estimateur du maximum de vraisemblance dans l'exemple donné plus haut.
Dans certains cas on connaît la distribution asymptotique de l'estimateur de la frontière. Dans le cas avec un *input* et un *output* Gijbels, Mammen, Park et Simar [63] dérivent la distribution asymptotique de l'estimateur de la frontière. Cet estimateur est super-convergent (il est $O_p(N^{\frac{2}{3}})$) et sa distribution asymptotique dépend de paramètres de nuisance liées au degré de *smoothness* de la fonction de production. Dans des cas plus généraux (avec plus d'un *input* ou plus d'un *output*), on sait que l'estimateur DEA converge vers la vraie frontière et on connaît la vitesse de convergence, voir par exemple Kneip, Park et Simar [81], mais les distributions asymptotiques sont inconnus à ce jour.

L'approche DEA fournit un estimateur convergent de la frontière de production (et donc des indices de croissance de la PGF et de ses décompositions) mais elle est purement statique. Färe, Grosskopf, Norris et Zhang [56] estiment un taux de croissance négatif de la frontière pour les États Unis et le Japon entre 1979 et 1980 et entre 1981 et 1982. Comment expliquer ces résultats ? En fait, puisque l'estimateur DEA sous estime la vraie frontière à distance finie, la frontière estimée correspond plus à une frontière de meilleure pratique (*best practice frontier*) qu'à une frontière technologique. Ainsi la régression de la frontière doit plutôt s'interpréter comme une baisse de l'efficience des économies *leader*, c'est à dire sur la frontière, que comme une régression technologique. Dans le cas de FGNZ [56], les États Unis déterminent la frontière technologique CRS et subissent une perte d'efficience sur les périodes 1979-1980 et 1981-1982.
Ceci suggère que la décomposition entre croissance de la technologie et crois-

sance de l'efficience proposée par FGNZ [56] n'est pas très satisfaisante dans un cadre dynamique. Dans cet exemple, il serait plus naturel de considérer la même frontière en 1979 et 1980 (ou en 1981 et 1982), c'est à dire de poser un progrès technologique nul, et de réduire l'efficience de l'économie *leader* (les États Unis). Cela reviendrait alors à contraindre le taux de croissance de la frontière à être positif ou nul. Cette contrainte ne change pas seulement le partage entre croissance de l'efficience et croissance de la technologie. En supposant une mémoire infinie des innovations on change aussi l'évaluation de l'indice de croissance de la productivité globale des facteurs.
On obtient très simplement une suite croissante d'ensembles de production estimés pour $t = 0, \dots, T$ en considérant à la date τ les observations des dates $t = 0, \dots, \tau$. Cette procédure est proposée par Diewert [38] et partiellement mise en œuvre par Henderson et Russell [79]. Dans la section suivante on ne se contente pas d'appliquer cette procédure, on modifie le processus générateur des données en supposant que l'ensemble de production à la date t est inclus dans l'ensemble de production à la date $t+1$. Ceci permet de montrer que contrairement à Henderson et Russell, il est important de considérer toute l'information disponible entre les dates $t = 0$ et T pour estimer la frontière en T.

2.2 Une présentation plus formelle

Dans cette section on décrit de façon formelle la méthode DEA, en insistant sur le modèle statistique justifiant cette approche.
A la date $t = 1, \dots, T$, pour chaque économie $n = 1, \dots, N$, on observe un vecteur $\mathbf{x}_{n,t} \equiv (x_{1,n,t}, x_{2,n,t}, \dots, x_{M,n,t}) \in \mathrm{X} \subseteq \mathbb{R}_+^M$ de M *inputs* utilisés dans le processus de production et un *output* $y_{n,t} \in \mathbb{R}_+$. A la date t et pour chaque économie n, on peut imaginer (sans l'observer) un niveau efficace du produit $y^*_{n,t} = \mathcal{F}_t(\mathbf{x}_{n,t})$, où la fonction $\mathcal{F}_t : \mathrm{X} \rightarrow R_+$ décrit la borne supérieure de l'ensemble de production à la date t. Ainsi le niveau efficace du produit est le niveau maximal du produit qu'une économie peut obtenir pour un panier d'*inputs* donné. La distance entre le produit observé et le produit efficace sera notée $\varepsilon_{n,t} \equiv \mathcal{F}_t(\mathbf{x}_{n,t}) - y_{n,t} = y^*_{n,t} - y_{n,t}$. Cette distance mesure la capacité d'une économie à mettre en œuvre la technologie mondiale à la date t.

Les notations et hypothèses suivantes permettent de caractériser la suite des ensembles de production. (Banker [11] formule les mêmes hypothèses, sauf l'hypothèse $\mathcal{H}6$, qui est liée à la croissance des ensembles de production, et l'hypothèse $\mathcal{H}7$ qui est sensiblement différente),

$\mathcal{H}1$ X est un sous ensemble convexe et compact de $\mathbb{R}_+^M$. L'ensemble

$$\mathrm{X}^* \equiv \left\{ \mathbf{x} \mid \mathbf{x} \geq \sum_{i=1}^{N} \lambda_i \mathbf{x}_i \,;\; \sum_{i=1}^{N} \lambda_i = 1 \,;\, \lambda_i \geq 0 \right\} \subseteq \mathrm{X}$$

désigne l'ensemble des combinaisons convexes des paniers d'*inputs*.

$\mathcal{H}2$ (MONOTONIE). Si $\mathbf{x}' \geq \mathbf{x}''$ alors $\mathcal{F}_t(\mathbf{x}') \geq \mathcal{F}_t(\mathbf{x}'')$.

$\mathcal{H}3$ (CONCAVITÉ). Si $0 \leq \lambda \leq 1$ alors $\lambda \mathcal{F}_t(\mathbf{x}') + (1-\lambda)\mathcal{F}_t(\mathbf{x}'') \leq \mathcal{F}_t(\lambda \mathbf{x}' + (1-\lambda)\mathbf{x}'')$.

$\mathcal{H}4$ (ENVELOPPEMENT). Pour chaque observation n à la date t, on a $y_{n,t} \leq \mathcal{F}_t(\mathbf{x}_{n,t})$.

$\mathcal{H}5$ (EXTRAPOLATION MINIMUM). A la date t, si une fonction $\widetilde{\mathcal{F}}_t(\bullet)$ satisfait les hypothèses [$\mathcal{H}2$] à [$\mathcal{H}4$] alors $\mathcal{F}_t(\mathbf{x}) \leq \widetilde{\mathcal{F}}_t(\mathbf{x})$.

$\mathcal{H}6$ (CROISSANCE). Pour tout $h \geq 0$ et panier $\mathbf{x} \in \mathrm{X}$ d'*inputs*, la suite de fonction de production satisfait $\mathcal{F}_t(\mathbf{x}) \leq \mathcal{F}_{t+h}(\mathbf{x})$.

Sous les hypothèses [$\mathcal{H}1$] à [$\mathcal{H}3$] la fonction de production $\mathcal{F}_t(\bullet)$ est continue sur le domaine défini par X[12]. L'hypothèse [$\mathcal{H}2$] sur la croissance de la fonction de production semble assez peu discutable. L'hypothèse [$\mathcal{H}3$], liée à la convexité de l'ensemble de production, peut sembler naturelle mais est bien plus discutable (surtout lorsque l'on considère une large coupe d'économies assez hétérogènes). Elle fera l'objet d'un test plus loin. L'hypothèse [$\mathcal{H}4$] contraint le produit à demeurer dans l'ensemble de production contemporain défini par la fonction $\mathcal{F}_\bullet(\bullet)$. On dira d'une économie qu'elle est efficiente si elle parvient à produire sur la borne supérieure de l'ensemble de production, c'est à dire si $y_{n,t} = \mathcal{F}_t(\mathbf{x}_{n,t})$. Une économie ne peut produire, à l'aide d'une manne divine, plus que la quantité définie par la fonction de production. Le point [$\mathcal{H}5$] pose qu'un ensemble convexe dont la borne supérieure satisfait les hypothèses [$\mathcal{H}2$]–[$\mathcal{H}4$] est supérieur ou égale à $\mathcal{F}_t(\bullet)$; dit autrement, $\mathcal{F}_t(\bullet)$

[12]Voir Rockafellar [125, section 10].

définit un ensemble de production *conservateur*. L'hypothèse [$\mathcal{H}6$] suppose que la suite d'ensembles de production est croissante, le processus générateur des données interdit les régressions technologiques.

Afin d'évaluer les propriétés de l'approche DEA en termes de convergence vers la vraie frontière, on doit définir la partie probabiliste du modèle. C'est l'objet de l'hypothèse suivante,

$\mathcal{H}7$ (EFFICIENCE). (i) Les écarts à la frontière $\varepsilon_{n,t}$ sont indépendamment et identiquement distribués. $f_\varepsilon(e)$ et $F_\varepsilon(\epsilon) = \int_{-\infty}^{\epsilon} f_\varepsilon(e)de$ représentent respectivement la densité marginale et la fonction de répartition associées à l'efficience. (ii) Conformément à l'hypothèse [$\mathcal{H}4$], on a $f_\varepsilon(e) = 0$ pour tout $e < 0$. (iii) Soit $h(\mathbf{x})$ la fonction de densité marginale associée au panier d'*inputs*. Pour tout $\mathbf{x} \in X \subseteq \mathbb{R}^M$, on a $\int_{\Omega_\delta(\mathbf{x})} h(\mathbf{x})d\mathbf{x} > 0$ où $\Omega_\delta(\mathbf{x})$ définit une boule ouverte dans X avec un rayon δ : $\Omega_\delta(\mathbf{x}) = \{x \in X; \|x - \mathbf{x}\| < \delta\}$. (iv) Pour tout η positif arbitrairement petit on a $\mathbb{P}\left(\varepsilon < \eta \;\middle|\; x \in \Omega_\delta(\mathbf{x})\right) > 0$ pour tout $\mathbf{x} \in X$.

Cette hypothèse est sensiblement différente de celle formulée par Banker [11], puisqu'ici on n'impose pas l'orthogonalité entre les paniers d'*inputs* et l'efficience[13]. Cette approche permet d'obtenir une estimation convergente de la frontière de production même si les inputs sont corrélés aux distances à la frontière, dès lors que la probabilité d'observer des réalisations du produit arbitrairement proches de la frontière est strictement positive.

L'estimateur statique

L'estimateur DEA de la fonction de production à la date t est défini par le problème de programmation linéaire suivant,

$$\begin{aligned}\widehat{\mathcal{F}}_t(\mathbf{x}) = \max\Bigg\{\mathbf{y} \in \mathbb{R}_+ \;\Bigg|\; \mathbf{y} = \sum_{n=1}^{N} \lambda_{n,t} y_{n,t} \;; \\ x_{j,t} \geq \sum_{n=1}^{N} \lambda_{n,t} x_{j,n,t} \text{ pour } j = 1, \ldots, M\Bigg\}\end{aligned} \tag{3.1}$$

[13]Comme dans Kneip, Park et Simar [81].

avec la suite de poids $\{\lambda_{n,t}\}_{n=1}^{N}$ vérifiant $\lambda_{n,t} \geq 0$ pour tout $n = 1, \ldots, N$ et $\sum_{n=1}^{N} \lambda_{n,t} = 1$[14].

Proposition 2.1. *Sous les hypothèses $\mathcal{H}1$–$\mathcal{H}5$ and $\mathcal{H}7$, l'estimateur statique [3.1] est convergent.*

Preuve : Banker [11], voir la section [5.3, en page 227]. ■

Banker [11] montre que, si l'efficience et les *inputs* sont orthogonaux, sous une condition supplémentaire l'estimateur DEA est l'estimateur du maximum de vraisemblance de la frontière technologique. L'estimateur DEA estime localement la frontière technologique en minimisant la distance $\varepsilon_{n,t} = y^*_{n,t} - y_{n,t}$ sous des contraintes liées à la forme de l'enveloppe linéaire par morceaux (concavité, croissance,...). Si la densité est décroissante[15], un estimateur du maximum de vraisemblance va aboutir à la même frontière, car dans ce cas en minimisant la distance à la frontière il maximise la vraisemblance.
Dans le cas avec deux facteurs de production, cet estimateur est $O_p(N^{-\frac{1}{2}})$ mais comme tout estimateur non paramétrique la vitesse de convergence décroît avec le nombre d'*inputs* (voir Simar et Wilson [139]).

Cette proposition établit la convergence de l'estimateur de type VRS vers la vraie frontière. On peut aussi montrer que l'estimateur de type NIRS tend vers la vraie frontière si celle-ci est de type NIRS ou CRS, et que l'estimateur de type CRS tend vers la vraie frontière si celle-ci est caractérisée par des rendements constants partout (voir Simar et Wilson [139]). L'estimateur de type VRS est donc robuste, dès lors que le vrai ensemble de production est convexe.

Évidemment , même si le processus générateur des données est augmenté avec l'hypothèse $\mathcal{H}6$ l'estimateur statique de la suite de frontières demeure convergent. Néanmoins, rien ne garantit que la suite des frontières estimées $\{\widehat{\mathcal{F}}_t(\mathbf{x}), t = 0, \ldots, T\}$ vérifie l'hypothèse $\mathcal{H}6$ à distance finie.

[14] Les contraintes sur ces pondérations impliquent des restrictions sur la nature des rendements d'échelle. Ici, les rendements d'échelle sont supposés non croissants et on ne contraint pas l'origine à être dans l'ensemble de production (il s'agit du cas VRS). On peut considérer d'autres types de contraintes. Par exemple, si les poids sont seulement contraints à être positifs on estime une frontière de production de type CRS.

[15] Si $0 \leq e \leq e'$ alors $f_\varepsilon(e) \geq f_\varepsilon(e')$.

L'estimateur dynamique

On utilise l'idée introduite plus haut pour obtenir une suite croissante d'ensembles de production estimés, conformément à l'hypothèse $\mathcal{H}6$. Sous les hypothèses $\mathcal{H}1$–$\mathcal{H}5$ et $\mathcal{H}6$ les nations sont capables de mélanger les innovations courantes et passées, dit autrement, les nations peuvent *convexifier* différentes technologies. Cette capacité est une condition nécessaire et suffisante pour obtenir une suite croissante d'ensembles de production. Cela suggère que si on peut estimer des ensembles de production de façon statique alors on peut construire une suite d'ensembles de production dynamiquement cohérente. L'algorithme suivant montre comment on peut procéder, en supposant que le monde est observé entre les dates 0 et T, pour *convexifier* le long des ensembles de production statiques passés,

Algorithme 2.2. *1. Estimer la frontière technologique statique à la date $t = 0$, $\widehat{\mathcal{F}}_0$, en utilisant l'estimateur défini par [3.1].*

2. *Sélectionner les observations sur la frontière estimée en 0 et les rajouter à l'échantillon des observations à la date $t = 1$.*
3. *Construire $\widehat{\mathcal{F}}_1^{\partial}$ sur la base de cet échantillon augmenté, toujours en utilisant [3.1].*
4. *Vérifier si les nations sélectionnées en $t = 0$ appartiennent à la frontière estimée $\widehat{\mathcal{F}}_1^{\partial}$. Si la réponse est négative, on élimine ces observations de l'échantillon.*
5. *Sélectionner les observations appartenant à la frontière en $t = 1$ et augmenter l'échantillon de la date $t = 2$ avec ces observations.*
6. *Continuer ainsi jusqu'en T.*

Cette procédure résulte en une suite croissante d'ensembles de production estimés, notée $\{\widehat{\mathcal{F}}_t^{\partial}(\mathbf{x}), t = 0, \ldots, T\}$, avec $\widehat{\mathcal{F}}_0^{\partial}(\mathbf{x}) = \widehat{\mathcal{F}}_0(\mathbf{x})$. Cette suite d'ensembles de production estimés dynamiquement cohérente est *conservatrice*, il n'est pas possible de construire une suite d'ensembles de production plus petits enveloppant les observations et vérifiant les hypothèses $\mathcal{H}1$–$\mathcal{H}6$.
Les caractéristiques de l'estimateur [3.1, en page 179] relativement aux rendements d'échelle sont préservées par cette procédure ; par exemple, si on utilise une spécification CRS l'algorithme [2.2] résultera en une suite croissante d'ensembles de production CRS.

Plutôt que d'utiliser cet algorithme on pourrait procéder en augmentant l'échantillon de la date τ avec toutes les observations passées. Cette alternative donne exactement les mêmes résultats, mais clairement elle est numériquement moins efficace (la dimension et le nombre des problèmes de programmation linéaire augmenteraient linéairement avec le temps).
Enfin, notons que cet estimateur sous-estime toujours la suite des vraies frontières. Néanmoins, il est possible d'établir la convergence de l'estimateur dynamique,

Proposition 2.3. *Sous les hypothèses $\mathcal{H}1$–$\mathcal{H}7$ l'estimateur dynamique défini par l'algorithme [2.2] est convergent.*

Preuve : Par construction on doit nécessairement vérifier les inégalités suivantes,

$$\widehat{\mathcal{F}}_t(\mathbf{x}) \leq \widehat{\mathcal{F}}_t^{\partial}(\mathbf{x}) \leq \mathcal{F}_t(\mathbf{x}) \ \forall t = 0, \dots, T$$

La convergence de l'estimateur statique implique la convergence de l'estimateur dynamique. ■

Notons $\mathcal{I}_t \equiv \bigcup_{\tau=0}^{t} \{(\mathbf{x}_{i,\tau}, y_{i,\tau})\}_{i=1}^{N}$ l'ensemble d'information à la date t et $\mathcal{I}_t^* \equiv \{(\mathbf{x}_{i,0}, y_{i,0})\} \cup \{(\mathbf{x}_{i,t}, y_{i,t})\}$ l'ensemble d'information restreint à la date t. Henderson et Russell considèrent l'ensemble d'information $\mathcal{I}_t^*$ plutôt que $\mathcal{I}_t$. Quel que soit l'ensemble d'information utilisé l'estimateur de la frontière à la date t sous-estime toujours la vraie frontière. Mais on vérifie nécessairement l'inégalité,

$$\widehat{\mathcal{F}}_{t,\mathcal{I}_t^*}(\mathbf{x}) \leq \widehat{\mathcal{F}}_{t,\mathcal{I}_t}(\mathbf{x})$$

En effet, par construction, on ne peut obtenir un ensemble de production plus petit en considérant plus d'information. Puisque ces deux estimateurs sous estiment la vraie frontière, on voit que l'estimateur basé sur $\mathcal{I}_t$ fournit une estimation moins biaisée pour tout N. Il est plus facile de se rapprocher de la vraie frontière en exploitant plus d'information.

2.3 Construction d'indices pour la PGF

En supposant que l'on connaisse la frontière technologique mondiale et donc les distances à cette frontière, on peut construire un indice de croissance de la productivité totale des facteurs. On adopte les notations suivantes.

$\mathcal{D}_t(\mathbf{x}_{n,t}, y_{n,t}) = \inf\{A_{n,t} | y_{n,t} \leq A_{n,t}\mathcal{F}_t(\mathbf{x}_{n,t})\}$ note la distance d'une économie n observée à la date t et la vraie frontière technologique à la date t. Cette distance vérifie $\mathcal{D}_t(\mathbf{x}_{n,t}, y_{n,t}) \in]0,1]$, si $\mathcal{D}_t(\mathbf{x}_{n,t}, y_{n,t}) = 1$ alors l'économie n se situe sur la vraie frontière technologique, sinon l'économie n'est pas efficiente. $D_t(\mathbf{x}_{n,t}, y_{n,t})$ note la distance entre l'économie n observée à la date t l'enveloppe conique[16] mondiale ($\mathcal{V}_t$) à la date t.

Même si la vraie frontière technologique mondiale (ou la vraie enveloppe conique) n'est pas observable, on a vu que l'on dispose d'un estimateur convergent. Les distances à la vraie frontière peuvent être estimées comme dans FGNZ [56]. L'estimateur de la distance d'une économie $\tilde{n}$ observée à la date t' relative à la frontière technologique à la date t (avec $t \gtreqless t'$) est défini par le programme suivant[17]

$$
\begin{aligned}
\left(\widehat{\mathcal{D}}_t(\mathbf{x}_{\tilde{n},t'}, \mathbf{y}_{\tilde{n},t'})\right)^{-1} &= \max_{A_{\tilde{n}}, \{\lambda_{n,t}\}_{n=1}^N} A_{\tilde{n}} \\
\underline{\text{s.t.}}\ A_{\tilde{n}} y_{\tilde{n},t'} &\leq \sum_{n=1}^{N} \lambda_{n,t} y_{n,t} \\
x_{\tilde{n},m,t'} &\geq \sum_{n=1}^{N} \lambda_{n,t} x_{n,m,t} \text{ for } m = 1, \ldots, M \\
\sum_{n=1}^{N} \lambda_{n,t} &= 1 \text{ et } \lambda_{n,t} \geq 0 \text{ for } n = 1, \ldots, N
\end{aligned}
$$

Pour estimer la distance à l'enveloppe conique on peut utiliser le même le problème de programmation linéaire en modifiant la restriction sur les poids λ_n (il s'agit du cas CRS où les poids vérifient seulement $\lambda_n \geq 0$ pour tout n). Pour estimer des distances cohérentes avec la croissance de la suite des ensembles de production, on modifie les problèmes de programmation linéaire conformément à l'algorithme [2.2].

On obtient les indices de croissance de la PGF comme des fonctionnelles des distances à l'enveloppe conique ($\mathcal{V}$) même si les rendements ne sont pas

[16] Qui correspond à la spécification CRS de la frontière technologique.

[17] On considère ici la spécification VRS car on sait que cet estimateur converge vers la vraie frontière quelle que soit la nature des rendements d'échelle de la vraie technologie.

constants partout. En procédant de cette façon on suit l'approche généralement employée dans cette littérature, par exemple Simar et Wilson [138] ou Ray et Desli [124]. Ces auteurs affirment que l'on doit considérer des distances à $\mathcal{V}$ pour calculer l'indice de croissance de la PGF. Dans le cas avec un *input* (x) et un *output* (y), la PGF est donnée par le ratio y/x quelle que soit la technologie convexe ($y = f(x)$). Ainsi l'indice de croissance de la PGF est mesuré sous l'hypothèse de rendements constants. Dans la littérature plus traditionnelle consacrée à la mesure de la croissance de la PGF, on formule aussi cette hypothèse. Par exemple Solow [141] suppose aussi que la technologie est à rendements constants. Mais cette hypothèse a surtout un avantage pratique car elle permet de recouvrir la part du capital dans le produit par complémentarité, voir Hulten [76]. Il n'est pas véritablement prouvé que l'hypothèse d'une frontière CRS, pour estimer les indices de croissance de la PGF, est d'autre justification que pratique. De la même façon dans la littérature DEA la considération d'une frontière CRS plutôt que VRS pour estimer les taux de croissance de la PGF n'est pas véritablement justifiée autrement que par des considérations pratiques. En effet, le taux de croissance de de la PGF sous l'hypothèse VRS n'est pas toujours identifiable (voir, par exemple, Ray et Desli [124]). On verra dans la section suivante que dans le cas de l'estimateur dynamique on peut toujours identifier le taux de croissance de la PGF. Ce n'est donc pas une convenance technique qui motivera notre choix de considérer une frontière CRS.

L'indice de croissance de la PGF entre les dates t et $t+1$ peut être calculé de deux façons : (i) en considérant que la technologie de référence est celle de la date t (ii) en considérant que la technologie de référence est celle de la date $t+1$. L'indice suivant repose sur une moyenne géométrique de ces deux cas,

$$M_t = \left(\frac{D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})}{D_t(\mathbf{x}_t, \mathbf{y}_t)} \times \frac{D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})}{D_{t+1}(\mathbf{x}_t, \mathbf{y}_t)} \right)^{\frac{1}{2}} \tag{3.2}$$

Chaque ratio mesure la croissance de la PGF comme la croissance de la distance à l'enveloppe conique entre les dates t et $t+1$. En considérant une moyenne géométrique, on obtient un indice indépendant de l'enveloppe conique de référence. Entre les dates t et $t+1$, la PGF s'accroît si et seulement si $M_t > 1$.

En suivant Simar et Wilson [138], on peut s'interroger sur les sources de la croissance de la PGF en décomposant l'indice (3.2) en un indice exprimant la croissance de l'efficience et un indice lié à la croissance de la frontière,

$$M_t = \underbrace{\Delta\mathrm{EFF}_t^{\mathcal{D}}\Delta\mathrm{ScaleEFF}_t}_{\Delta\mathrm{EFF}_t} \times \underbrace{\Delta\mathrm{TECH}_t^{\mathcal{D}}\Delta\mathrm{ScaleTECH}_t}_{\Delta\mathrm{TECH}_t} \tag{3.3}$$

L'indice $\Delta\mathrm{EFF}_t$ mesure le facteur de croissance de l'efficience en comparant les distances aux enveloppes contemporaines aux dates $t+1$ et t,

$$\Delta\mathrm{EFF}_t = \frac{D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})}{D_t(\mathbf{x}_t, \mathbf{y}_t)} \tag{3.4}$$

Mais cet indice ne rend pas compte de façon satisfaisante de la croissance de l'efficience si les rendements d'échelle ne sont pas constants partout. Cette limite suggère une seconde décomposition. $\Delta\mathrm{EFF}_t^{\mathcal{D}}$ correspond au facteur de croissance sous la spécification VRS de la technologie (on écrit la définition formelle de cet indice en remplaçant D par $\mathcal{D}$ dans l'équation (3.4)) et $\Delta\mathrm{ScaleEFF}_t$ mesure la déviation à la spécification CRS relativement à l'efficience. L'indice $\Delta\mathrm{TECH}_t$ mesure la croissance de la technologie en comparant les niveaux de l'enveloppe conique aux dates $t+1$ et t. Évidemment, cette mesure doit dépendre du panier d'*inputs* caractérisant l'économie considérée à chaque date. Afin d'éliminer cette dépendance on utilise à nouveau une moyenne géométrique,

$$\Delta\mathrm{TECH}_t = \left(\frac{D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})}{D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})} \cdot \frac{D_t(\mathbf{x}_t, \mathbf{y}_t)}{D_{t+1}(\mathbf{x}_t, \mathbf{y}_t)}\right)^{\frac{1}{2}} \tag{3.5}$$

Cet indice ne mesure pas correctement la croissance de la technologie si les rendements ne sont pas constants partout. Cela motive, à nouveau, une seconde décomposition. $\Delta\mathrm{TECH}_t^{\mathcal{D}}$ mesure la croissance de la technologie pour la spécification VRS de la technologie et $\Delta\mathrm{ScaleTECH}_t$ mesure la déviation à la spécification CRS relativement à la technologie. Notons que, contrairement à l'indice d'échelle précédent, cet indice ne dépend pas du panier d'*inputs* considéré.

Il convient de noter que les décompositions, liées à la nature des rendements, proposées ici sont discutables. Nous avons suivi les recommendations de Simar et Wilson [138], mais on peut trouver d'autres décompositions dans la littérature. On peut se reporter à Grosskopf [67] pour une discussion. Dans la section suivante on suppose que les rendements sont constants et on n'insiste pas sur ces décompositions.

3 Comptabilité de la croissance

On applique l'approche présentée dans la section [2] à un échantillon de 57 économies observées entre 1965 et 1990 (voir la section [5.1, en page 222]). Les tables [3.1, en page 233] and [3.2, en page 234] donnent une présentation synthétique de cet échantillon[18].

Le tableau [3.3, en page 235] présente les estimations obtenues pour le taux de croissance annuel moyen de la PGF sous l'hypothèse CRS (et ses décompositions, obtenues sous l'hypothèse VRS) en utilisant l'estimateur statique[19]. Le tableau [3.4, en page 236] reporte les estimations du taux de croissance annuel moyen de la PGF (et ses décompositions) en utilisant l'estimateur dynamique. A quatre exceptions près (l'Équateur, le Paraguay, les Pays Bas et la Suède) les signes du taux de croissance de la PGF sont identiques dans les deux tables, mais les ordres de grandeur des estimations diffèrent sensiblement.

Il est intéressant de comparer les taux de croissance de l'efficience pour les deux types d'estimateur, voir les tableaux [3.5, en page 237] et [3.6]. Dans le cas de l'estimateur statique, le taux de croissance moyen de l'efficience est positif alors qu'avec l'estimateur dynamique il est négatif. En moyenne, les nations s'éloignent de la frontière technologique mondiale lorsque l'on exclut la régression technologique. On obtient donc un résultat différent de Kumar et Russell [84] qui observent une tendance moyenne des économies à rejoindre la frontière technologique. Cette différence s'explique très simplement. En effet, on sait que la frontière estimée avec l'estimateur dynamique domine la frontière estimée par la frontière statique (voir la section [2.2]). Ainsi les distances estimées à la frontière technologique en 1990 sont plus importantes avec l'estimateur dynamique. Ceci résulte en un taux de croissance de l'efficience plus faible lorsque l'on considère l'estimateur dynamique. Si, comme le suppose le processus générateur des données envisagé dans la section [2.2],

[18]Nous pourrions utiliser des données plus désagrégées pour le stock de capital, mais cela altérerait considérablement la vitesse de convergence des estimateurs utilisés (*Curse of dimensionality*). Rappelons que dans le cas avec deux *inputs* l'estimateur VRS statique de la frontière technologique est $O_p(N^{-\frac{1}{2}})$.

[19]On peut calculer l'estimateur VRS du taux de croissance de la PGF à partir des décompositions présentées dans le tableau : $\Delta TFP^{\mathcal{D}} = (1+\Delta EFF^{\mathcal{D}})(1+\Delta TECH^{\mathcal{D}})-1$.

le progrès technologique est toujours positif ou nul, on doit accorder plus de crédit au taux de croissance négatif de l'efficience obtenu dans ce chapitre qu'au taux de croissance positif obtenu par Kumar et Russell [84], car ces auteurs sur estiment nécessairement le taux de croissance de l'efficience. Henderson et Russell, alors qu'ils excluent la régression technologique, concluent aussi à un rapprochement vers la frontière[20]. Ce résultat s'explique par le fait qu'ils ne considèrent pas les observations entre 1965 et 1990. En effet, ils obtiennent ainsi un estimateur qui sous estime plus la vraie frontière en 1990 que notre estimateur. Notre frontière estimée domine leur frontière estimée, ce qui explique, une nouvelle fois, la différence en termes de taux de croissance moyen de l'efficience. La figure [3.16, en page 244] montre dans quelles régions le biais de Henderson et Russell [79] est plus important que le nôtre.

Enfin notons que dans le cas de l'estimateur statique on ne parvient pas à estimer le taux de croissance de la technologie pour quatre économies, alors que dans le cas de l'estimateur dynamique on ne rencontre jamais ce problème. Simar et Wilson [138] sont eux aussi confrontés à ce problème (pour l'Irlande). Lorsque l'on veut estimer la distance à une frontière non contemporaine, il arrive que le problème de programmation linéaire n'ait pas de solution, dans le cas d'une spécification VRS de la technologie. La figure [3.9, en page 231] représente un tel cas. Le point rouge est une observation de la date t, on voit bien qu'il n'est pas possible de définir sa distance par rapport à la frontière de production future. Ce cas ne peut survenir que dans le cas d'une spécification VRS, lorsque le coût d'accès à la technologie mondiale augmente. Il ne peut survenir lorsque l'on utilise l'estimateur dynamique. Puisque dans ce cas une nation peut toujours utiliser une technologie passée, il n'y a pas de raison pour que le coût d'accès à la technologie mondiale augmente (il ne peut que diminuer). Dès lors on peut toujours estimer l'indice de croissance de la frontière VRS.

3.1 La frontière technologique mondiale

La figure [3.11] représente les estimations de la frontière technologique mondiale pour les années t = 1965, 1970, 1975, 1980, 1985 et 1990, dans

[20]Lorsqu'ils ne considèrent pas, comme dans ce chapitre, le capital humain comme un *input* de la fonction de production mondiale.

le plan capital par travailleur – produit par travailleur. Pour estimer ces frontières, on a supposé que la frontière technologique mondiale est à rendements constants (ainsi, dans le plan considéré la technologie est de type NIRS). Cette hypothèse est naturelle dans le sens où il n'est pas possible d'écrire un modèle de croissance exhibant un état stationnaire non trivial avec des rendements globalement décroissants.

Les frontières technologiques sont déterminées par plus d'une nation. Sur la période 1965–1990, six économies rencontrent la suite de frontières estimées (SLE, USA, ARG, PRY, HKG et LUX), voir le tableau [3.7, en page 238]). Durant deux années consécutives (1982 et 1983) il n'y a pas d'économie sur la frontière estimée, les économies *leader* expérimentent alors un recul de l'efficience. Un estimateur statique de la suite de frontières technologiques conclurait à une régression technologique.

On pourrait s'étonner de la présence d'économies comme la Sierra Leone ou le Paraguay sur la même frontière de production que les États Unis. Si ces deux économies étaient capables de mobiliser autant de capital par travailleur que les États Unis, elles devraient être capables d'atteindre le même niveau de production par travailleur. On peut limiter ce raisonnement dans deux directions. En premier lieu, ce raisonnement suppose implicitement que si une économie pauvre efficace devient plus intensive en capital alors elle demeure toujours aussi efficace. Cet argument n'est recevable que si l'efficacité technique d'une nation ne dépend pas de sa position dans la distribution du capital par travailleur. Le processus générateur des données décrit plus haut ne formule pas cette hypothèse forte. Si on pense la technologie mondiale, qui n'est qu'une vue de l'esprit, comme une combinaison convexe de différentes technologies, un niveau du capital par travailleur correspond à une technologie particulière. Ainsi, un changement du stock de capital par travailleur est accompagné d'un changement de la technologie à laquelle est confrontée une économie. Il n'y a aucune raison de penser qu'une nation, formée sur une technologie peu intensive en capital puisse sans problème s'adapter à une technologie plus intensive en capital. Deuxièmement, il faut rappeler que l'approche DEA sous-estime la vraie frontière. Ainsi, il y a peu de chance pour qu'une économie sur la frontière estimée soit sur la vraie frontière et qu'elle soit donc réellement efficace. Notons enfin que la présence de ces deux nations sur la même frontière que les États Unis peut remettre en question l'hypothèse d'un unique ensemble de production convexe.

On peut observer que la déformation de l'ensemble de production n'est pas homothétique. L'hétérogénéité des taux de croissance de la frontière est relativement importante. Le taux de croissance moyen dans la coupe est inférieur à 1% (approximativement 0.6%). La distribution des taux de croissance est asymétrique (voir la médiane dans le tableau [3.6, en page 237]) et on peut voir que le taux de croissance est d'autant plus important que l'on se situe aux extrémités de la distribution du capital par travailleur (voir les tableaux [3.1] et [3.4], et la figure [3.11]). Le taux de croissance de la technologie est d'espérance nulle dans un voisinage du capital par travailleur moyen. On trouve donc comme Kumar et Russell des indices contre l'idée d'un progrès technique neutre. Mais chez ces deux auteurs le taux de croissance de la technologie est d'autant plus fort que l'intensité capitalistique est importante.

3.2 Sources de la croissance

Nous abordons maintenant l'objet de ce chapitre, quelles sont les contributions respectives des mécanismes néo-classique (associé à l'accumulation du capital) et technologique (identifié par la productivité globale des facteurs) à la croissance du produit par travailleur ?

Globalement, la croissance de la PGF ne semble pas fortement liée à la croissance du produit par travailleur. Le taux de croissance moyen dans la coupe du produit par travailleur entre 1965 et 1990 est approximativement de deux pour cent alors que le taux de croissance annuel moyen de la PGF est autour de 0,4%. Ainsi la contribution moyenne de la PGF à la croissance du produit par travailleur est de l'ordre de un cinquième.

Si la contribution de la PGF semble marginale en moyenne, un regard plus proche des données montre que cette contribution est variable dans la coupe. Alors que la croissance de la PGF égalise la croissance du produit par travailleur pour des pays comme la Norvège et la Suisse, sa contribution est imperceptible pour une économie comme la Turquie (voir le tableau [3.6, en page 237]). Cette hétérogénéité suggère une approche locale, car rien ne nous laisse penser que la contribution de la PGF devrait être constante dans la coupe.

La figure du haut dans [3.12, en page 240], qui peut être vue comme une version continue de la figure 1 dans Easterly et Levine [48, page 40], montre le taux de croissance de la PGF conditionné par le taux de croissance du produit par travailleur. Un point $(\Delta y, \Delta PGF)$ le long de cette courbe donne le taux de croissance moyen de la PGF pour des économies similaires en termes de croissance de la PGF. La figure du bas montre la contribution de l'accumulation du capital au taux de croissance du produit par travailleur.
Il apparaît clairement qu'en utilisant un estimateur de l'espérance linéaire de la croissance de la PGF conditionnelle à la croissance du produit par travailleur plutôt que d'un estimateur de l'espérance non linéaire (comme on le fait en considérant les moyennes empiriques de la croissance de la PGF et du produit par travailleur dans le tableau [3.6, en page 237]), on perd une information importante. La figure [3.12] révèle que la croissance de la PGF est marginale dans le cas des miracles économiques. Ce résultat est cohérent avec les estimations de Young [154] sur les performances des pays du sud-est asiatique. A l'inverse, la croissance de la PGF apparaît fortement liée aux désastres économiques. Pour ces nations, la contribution de l'accumulation du capital est proche de zéro et la dynamique du produit se réduit à la dynamique de la productivité globale des facteurs. Il convient cependant de noter que ces observations ne reposent que sur le premier moment des distributions. Même dans le cas des miracles économiques, on observe une hétérogénéité relativement importante. Par exemple, la contribution de la croissance de la PGF à la croissance du produit par travailleur est de 18,4% pour la Corée du Sud contre 65,8% pour Hong Kong. Cela suggère que nous devrions tenir compte de tous les moments. La figure [3.13, en page 241] est construite dans cette perspective. Elle confronte la densité estimée du taux de croissance du produit par travailleur à la densité estimée du produit par travailleur induit par l'accumulation du capital (ou la croissance de la PGF). Clairement, l'accumulation du capital est le mécanisme le plus important dans les performances des nations en termes de croissance. Li [91] propose une statistique unilatérale pour tester l'égalité de deux densités inconnues[21].

[21]Cette statistique est asymptotiquement gaussienne. Même si, pour notre échantillon,

Pour le cadre du haut (resp. le cadre du bas) cette statistique vaut $T_1 = 0,566$ (resp. $T_1 = 7,377$). Ces statistiques ne réfutent pas l'idée que la dynamique du produit par travailleur est essentiellement liée à l'accumulation du capital.

cette approximation asymptotique peut se révéler hasardeuse, la statistique donne une idée de la distance entre les distributions comparées.

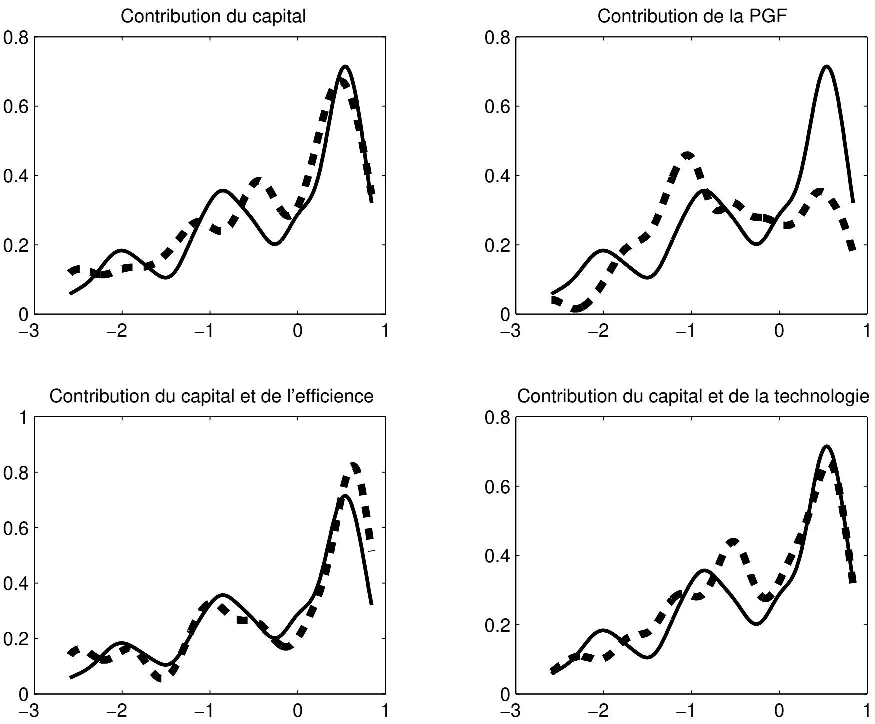

FIG. 3.1 – **Contributions à la distribution du produit par travailleur relatif en 1990**. Dans les quatre cadrans la courbe pleine représente un estimateur à noyau de la densité du produit par travailleur relatif en 1990 (le paramètre de lissage est choisi par une procédure de *cross-validation*). La courbe en tirets dans le cadre de gauche en haut représente une estimation de la distribution du produit par travailleur relatif que l'on obtiendrait en 1990 sans croissance de la PGF. La courbe en tirets dans le cadre de droite en haut représente une estimation de la distribution du produit par travailleur relatif que nous obtiendrions en 1990 sans croissance du capital par travailleur. La courbe en tirets dans le cadre de gauche en bas représente une estimation de la distribution du produit par travailleur en 1990 sans croissance de la technologie. La courbe en tirets dans le dernier cadre représente une estimation du produit par travailleur relatif en 1990 sans croissance de l'efficience.

Une perspective alternative est de rechercher les contributions respectives de la PGF, de l'accumulation du capital, de la croissance de l'efficience et du progrès technologique à la distribution du produit par travailleur en 1990. Dans quelle mesure l'accumulation du capital est-elle capable d'expliquer la forme multi-modale du produit par travailleur en 1990 ? La figure [3.1] est conçue pour répondre à cette question. En utilisant le taux de croissance estimé de la PGF, de l'efficience et de la technologie, on simule la distribution du produit par travailleur en 1990 en posant à zéro la croissance de la PGF, la contribution de l'accumulation du capital, la croissance technologique et la croissance de l'efficience entre 1965 et 1990. En comparant ces distributions simulées avec la distribution estimée du produit par travailleur en 1990 on obtient une idée sur les contributions respectives de l'accumulation du capital, de la croissance de la PGF, de la croissance de l'efficience et de la croissance de la technologie aux inégalités de richesse en 1990. On voit que la contribution de l'accumulation du capital physique est essentielle. Ce mécanisme suffit pour reproduire une distribution multi-modale du produit par travailleur. On note cependant que si ce facteur reproduit fidèlement le haut de la distribution du produit par travailleur, il a moins de réussite dans le bas de la distribution : les modes des deux distributions ne correspondent pas. La masse de probabilité de la distribution du produit par travailleur relatif en 1990 simulé sans croissance de la PGF est plus à droite. Si on augmente la dynamique induite par l'accumulation du capital avec la dynamique de l'efficience on reproduit mieux le bas de la distribution du produit par travailleur en 1990. Cela suggère que si la dynamique d'efficience, *ie* la capacité des nations à exploiter la frontière technologique mondiale, a un rôle dans la distribution du produit par travailleur en 1990, c'est surtout en appauvrissant les économies dans le bas de de la distribution[22]. Notons que le pouvoir explicatif de la PGF est bien plus faible et que la croissance de la frontière n'aide pas à se rapprocher de la distribution du produit par travailleur observé.

Toutes ces observations sont cohérentes : (*i*) l'accumulation du capital semble être le mécanisme principal, au moins pour les nations dans le haut de la distribution du produit par travailleur en 1990 et (*ii*) la dynamique de la productivité globale des facteurs, essentiellement la dynamique de l'efficience, joue un rôle plus important dans le bas de la distribution.

[22]Notons que nous présentons des estimations de la densité du produit par travailleur en 1990 bien moins lisses que Kumar et Russell [84] ou Henderson et Russell [79]. Ces auteurs ont vraisemblablement choisi un paramètre de lissage trop important (rappelons qu'une augmentation du paramètre de lissage réduit la variance mais augmente le biais), ils ne précisent pas quel critère ils ont utilisé pour déterminer ce paramètre. Ils éliminent ainsi des informations importantes et ne peuvent réellement commenter le rôle des différentes sources de la croissance.

3.3 Contributions à la convergence

On s'intéresse ici plus spécifiquement à la mobilité des nations à l'intérieur de la distribution du produit par travailleur. Un fait stylisé bien connu est que ces 57 nations divergent. Les figures [3.2] et [3.14, en page 242] considèrent deux sources de convergence, l'effet de *capital deepening* (modèle de Solow-Swan) et le *catch-up* technologique (modèle d'Abramovitz), et peuvent aider à mettre en œuvre une sorte de comptabilité qualitative de la divergence.

Le cadre nord-ouest de la figure [3.2] représente une estimation de l'espérance non linéaire du taux de croissance du capital par travailleur conditionnelle au capital par travailleur relatif initial. La relation entre le taux de croissance du capital intensif et son niveau initial est non linéaire. Ce graphique montre que la dynamique du capital est caractérisée par deux régimes de convergence. Si le stock de capital par travailleur relatif initial appartient à l'ensemble $[-4, -2] \cup [0, 2]$ la relation est décroissante ; ainsi dans cette région on observe une convergence des nations en termes de stock de capital par travailleur. Si le stock de capital relatif par travailleur initial est dans l'intervalle [-2,0] la relation est croissante, les économies divergent car un niveau initial plus important implique en moyenne un taux de croissance plus important. Ce découpage est cohérent avec notre observation précédente, la dynamique du capital par travailleur peut produire le caractère multi-modal de la distribution. Ceci suggère l'existence de deux *clubs* de convergence. Le cadre sud-est de la figure [3.2] représente une estimation de l'espérance non linéaire de la croissance de l'efficience conditionnelle au niveau initial du stock de capital par travailleur relatif. Cette relation est monotone croissante et passe par l'origine. L'efficience d'une économie s'accroît en espérance si et seulement si son niveau initial de stock de capital par travailleur relatif est positif. Pour qu'une économie se rapproche de la frontière technologique, il faut que sa dotation en capital soit suffisamment importante. Une nation initialement faiblement intensive en capital s'éloigne de la frontière technologique en espérance. Le cadre sud-ouest de la figure [3.2] représente une estimation de l'espérance non linéaire de la croissance de l'efficience conditionnelle au niveau initial de l'efficience. La relation est en forme de U inversé. Cette relation est décroissante si et seulement si le niveau initial de l'efficience est supérieur à 0,6. Dans cette zone, néanmoins, le taux de croissance espéré de

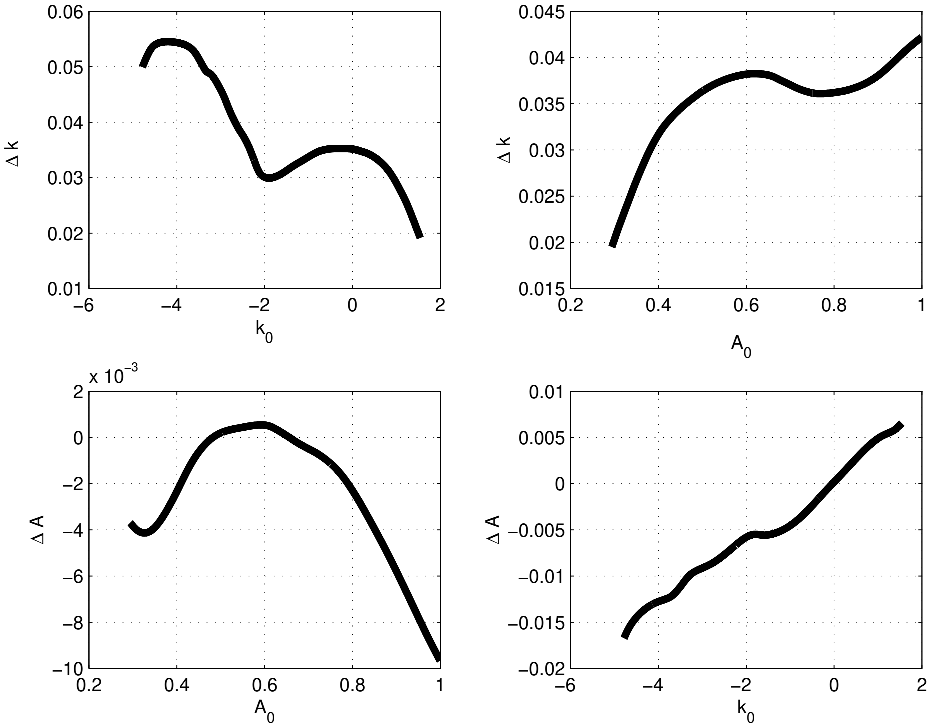

FIG. 3.2 – **Sources de la convergence (I)**. Toutes ces courbes sont obtenues par des régressions non linéaires (*local polynomial regression*).

l'efficience est généralement négatif. Ainsi une économie s'éloigne d'autant plus rapidement de la frontière qu'elle en est proche. Si l'efficience initiale est inférieure à 0,6, la relation est croissante et le taux de croissance espéré de l'efficience est généralement négatif. On a alors une divergence technologique double : (i) les économies s'éloignent de la frontière (ii) elles s'en éloignent d'autant plus vite qu'elles en sont éloignées.

Les cadres nord-est et sud-est de la figure [3.2] suggèrent que l'on ne peut considérer l'effet de *capital deepening* et *catch-up* technologique de façon indépendante. Même un simple modèle de croissance optimale néo-classique augmenté avec un mécanisme de *catch-up* technologique[23] prédit une relation positive entre la croissance du capital par travailleur et l'efficience initiale (cadre nord-est). Le cadre sud-est pourrait être expliqué par un problème

[23]Voir le chapitre [1, en page 41].

d'adoption de la meilleure technologie : une économie doit être assez intensive en capital pour bénéficier de la diffusion de la technologie *leader*.
La figure [3.14, en page 242] représente une régression non paramétrique de la croissance du produit par travailleur relatif aux États Unis sur les conditions initiales de l'efficience et du stock de capital par travailleur relatif. Clairement, une nation bénéficie plus facilement de son retard technologique si sa dotation initiale en capital par travailleur est proche de celle des États Unis. La surface représentée dans cette figure est en forme de bol inversé[24], la zone culminante se situant dans la région où le capital par travailleur initial est proche de celui des états unis et où le retard technologique est d'importance moyenne (autour de 0,6). Généralement, les économies proches de la frontière en 1965 ont tendance à s'éloigner des États Unis, comme les économies éloignées de la frontière et peu intensives en capital.

Les résultats reportés dans ce chapitre contredisent les conclusions des exercices de *growth accounting* mis en œuvre par Easterly et Levine [48] qui obtiennent une contribution de la PGF à la croissance supérieure à 50%. Comment une telle différence peut-elle s'expliquer ? Easterly et Levine considèrent un échantillon de 64 nations extrait des *Penn World Tables* 5.6, sur une période plus courte (1980–1992). Ils supposent que la technologie est Cobb-Douglas et que la part du capital est 0,4 ; ils obtiennent alors une part de la PGF dans la croissance du produit par travailleur de 60%. La différence avec leurs résultats peut s'expliquer par la période considérée[25]. Dans la Figure [3.15, en page 243] on reprend l'exercice proposé dans la figure [3.13, en page 241] sur deux sous-périodes (1965–1980 et 1980–1990). Durant la seconde période, qui est proche de celle considérée par Easterly et Levine, la contribution de la croissance de la PGF à la croissance du produit par travailleur apparaît bien plus évidente. Entre 1980 et 1990, la distribution du taux de croissance annuel moyen du produit par travailleur est multi-modale, et la dynamique de la productivité globale des facteurs parvient à reproduire cette forme (*ie*, les deux modes les plus importants de la distribution). Sur la première sous période, on retrouve une contribution faible de la croissance

[24]Sans considérer le rebond dans le fond de la figure pour un niveau important du capital par travailleur initial relatif, où il n'y a pas d'observations.

[25]Sans compter que les deux approches sont radicalement différentes puisqu'il n'y a pas d'estimation dans le cas d'Easterly et Levine.

de la PGF. Cette hétérogénéité chronologique de la contribution de la PGF à la croissance du produit par travailleur peut s'expliquer par la proximité des économies de leurs états stationnaires respectifs dans les années 80. Entre 1965 et 1980 le processus néo-classique de convergence, dirigé par l'accumulation du capital, domine. Après 1980, une fois que l'effet du *capital deepening* s'épuise, la dynamique du produit par travailleur est essentiellement déterminée par la dynamique de la PGF. Ces observations sont cohérentes avec les faits empiriques données dans l'introduction générale, où on a vu la stabilité de la distribution du produit par tête à partir des années 70-80. Ainsi, Easterly et Levine biaisent leurs résultats en faveur de l'explication technologique de la croissance en considérant la période 1980-1992.

3.4 Expliquer les distances à la frontière

Dans la section précédente on a vu que la dynamique d'efficience (après l'accumulation du capital) joue un rôle non négligeable pour expliquer les performances des nations en termes de croissance et niveau. La capacité (ou l'incapacité) d'une économie à exploiter la frontière mondiale explique en grande partie les contre-performances des économies dans le bas de la distribution des richesses. La figure [3.3] représente des estimateurs à noyau de la distribution des distances à la frontière en 1965, 1970, 1975, 1980, 1985 et 1990. On observe que la masse de probabilité se déplace vers la gauche, *ie* les économies tendent à s'éloigner de la frontière, et la distribution de l'efficience semble tendre vers une distribution bi-modale. Cette observation est conforme avec la relation en forme de U inversé entre la croissance de l'efficience et le niveau initial de l'efficience (le cadre sud-ouest de la figure [3.2, en page 196]).

Nous considérons ici le rôle de l'éducation dans l'hétérogénéité des distances à la frontière. La figure [3.4] est conçue pour évaluer l'effet de l'éducation sur l'efficience. Il s'agit d'une *stochastic kernel* (une version continue d'une matrice de transition, voir Quah [123]) reliant la distribution de l'efficience (pour les années 1965, 1970, 1975, 1980, 1985 and 1990[26]) et la distribution

[26]On suppose que la relation entre capital humain et efficience est invariante dans la chronologie.

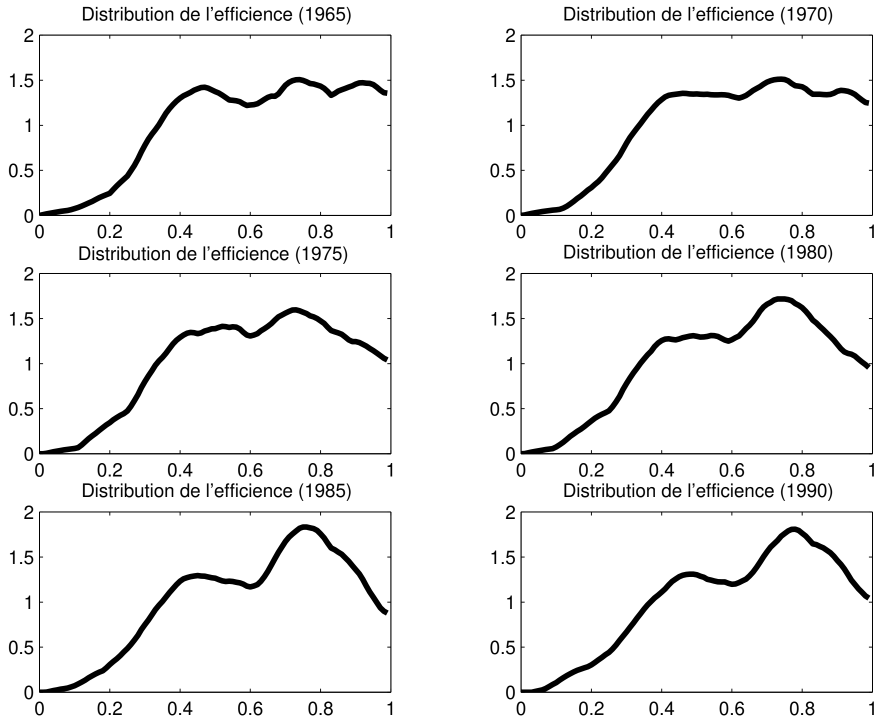

Fig. 3.3 – **Distribution de l'efficience**. Ces estimateurs à noyau de la densité de l'efficience sont obtenues avec une réflexion sur les bornes, comme décrit dans Silverman [132, pages 29-32]. Le paramètre de lissage commun est 0,17. Une procédure de *cross-validation* (sans considérer les économies efficientes, voir Silverman [132, pages 51-52]) donnerait un choix du paramètre de lissage généralement légèrement supérieur (de l'ordre de 0,2), et le caractère bi-modale de la distribution apparaîtrait moins évident.

de l'efficience conditionnée par le stock de capital humain[27]. On observe un remarquable changement de la distribution inconditionnelle à la distribution conditionnelle. Si l'éducation n'explique pas l'hétérogénéité des distances à la

[27] L'efficience conditionnée par le stock de capital humain est le niveau d'efficience moyen des nations qui sont proches en termes de capital humain. Soient $\{D_i\}_{i=1,\dots,N}$ les niveaux d'efficience et $\{H_i\}_{i=1,\dots,N}$ les niveaux d'éducation. Le niveau de l'efficience conditionnelle au niveau de l'éducation est noté $\{\overline{D}_i\}_{i=1,\dots,N}$. $\overline{D}_j$ est l'efficience espérée des économies dont le niveau d'éducation est proche de H_j. On utilise une régression non paramétrique (Nadaraya-Watson) pour obtenir cette quantité. Les données sur le capital humain sont extraites de Barro et Lee [13] (on considère la variable *secondary school complete, percentage in total population*, c'est à dire le pourcentage dans la population totale des individus ayant complété leurs études secondaires).

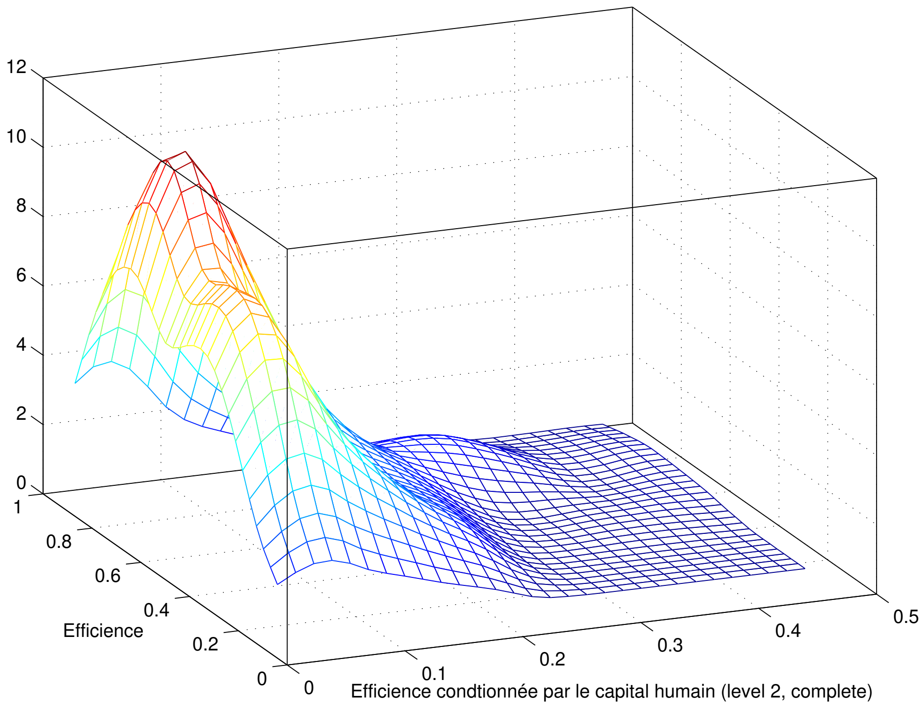

Fig. 3.4 – **Expliquer la distribution de l'efficience**.

frontière alors la masse de probabilité doit se répartir le long de la première bissectrice, ainsi la distribution conditionnelle n'est pas différente de la distribution inconditionnelle. Dans le cas qui nous intéresse, si le niveau de capital humain ne semble pas capable d'expliquer la distance moyenne à la frontière, L'essentiel de la masse de probabilité de la distribution bivariée s'étend parallèlement à l'axe correspondant à l'efficience inconditionnelle. Cela veut dire que l'hétérogénéité des niveaux d'éducation explique l'hétérogénéité des distances à la frontière, c'est à dire des propensions à mettre en œuvre la technologie mondiale.

Évidemment il serait intéressant d'envisager d'autres variables. Peut être pourrions-nous trouver des variables capables d'expliquer, plus particulièrement, le niveau moyen de l'efficience. Notons enfin que ce dernier résultat tend à souligner le rôle non négligeable du capital humain dans la croissance. On pourrait regretter de ne pas voir cette variable apparaître comme un *input* de la fonction de production mondiale, de la même façon que chez

Henderson et Russell [79]. En adoptant une stratégie en deux étapes, on a implicitement opté pour un cadre théorique très particulier. Ici le capital humain n'intervient qu'au travers du terme d'efficience, un peu comme dans Nelson et Phelps [102]. Mais chez ces auteurs le capital humain affecte directement le taux de croissance de l'efficience et non pas le niveau. Ici on montre que l'hétérogénéité sur le stock de capital humain explique l'hétérogénéité des distances à la frontière. Notons néanmoins que l'introduction du capital humain dans la fonction de production chez Henderson et Russell [79] est aussi très particulière. Ils choisissent de faire intervenir le capital humain en ajustant le nombre de travailleur de la qualité de l'éducation. Il s'agit toujours d'une fonction de production à deux facteurs, l'éducation intervient comme un terme croisé avec le nombre de travailleurs. D'une certaine façon, ce choix contredit l'approche non paramétrique proposée par les deux auteurs. En toute généralité on devrait voir apparaître le stock de capital humain comme un troisième *input* de la technologie. L'intervention de l'éducation sous la forme d'un terme croisé est une hypothèse paramétrique sur la forme de la technologie. Dans un cadre totalement différent, car ils supposent que la technologie est Cobb-Douglas, Benhabib et Spiegel [18] montrent que le capital humain n'améliore pas la prédiction du taux de croissance du produit s'il apparaît comme un *input* de la technologie. Ces deux auteurs suivent alors la spécification de Nelson et Phelps et montrent que le rôle du capital humain paraît ainsi plus évident. Ces différentes spécifications pourraient faire l'objet de tests dans le cadre de l'approche DEA.

3.5 limites et prolongements

Pour résumer, la technologie n'apparaît pas être le moteur essentiel de la croissance, la dynamique de la PGF ne peut pas rendre compte de l'hétérogénéité des dynamiques du produit par travailleur. Easterly and Levine [48] écrivent,

> "*Economists should devote more effort towards modelling and quantifying total factor productivity*"

En utilisant une approche plus générale que celle employée par ces deux auteurs, on ne trouve pas d'évidence empirique étayant leur message selon lequel la technologie, et non l'accumulation de facteurs, est l'explication majeure de la diversité des performances nationales. Néanmoins, on montre que la dynamique de la PGF, essentiellement la dynamique d'efficience, enrichit le phénomène de divergence caractérisant cet échantillon. Les nations dans le bas de la distribution des richesses s'éloignent de la frontière technologique mondiale, alors que les économies dans le haut de la distribution des richesses s'en rapprochent en moyenne.

Un des résultats originaux de cette section est de montrer que globalement on ne trouve pas d'éléments en faveur de l'hypothèse du *catch-up* technologique (dans le sens d'Abramovitz [2]). Une comparaison avec Kumar et Russell [84] et Henderson et Russell [79] montre que cette conclusion résulte de l'hypothèse sur la croissance de la suite des ensembles de production. Si les données sont effectivement générées par un modèle assurant un progrès technologique positif ou nul, alors les résultats reportés dans cette section sont moins biaisés que les résultats reportés par ces auteurs.
La décomposition de l'indice de croissance de la PGF en la croissance de l'efficience et la croissance de la frontière donne un instrument pour évaluer la capacité des nations à adopter la meilleure pratique. Avec cette mesure à disposition on peut interroger les données sur les sources de l'hétérogénéité relative à la propension d'une économie à rattraper la frontière technologique. Dans la section [3.4], on a montré que le niveau de l'éducation est un bon candidat pour expliquer cette hétérogénéité.

Dans cette section nous n'avons pas proposé de tests formels. On sait que l'estimateur DEA converge vers la vraie frontière et on connaît la vitesse de convergence de cet estimateur. Sa distribution asymptotique est plus difficile à établir mais on peut (sous des hypothèses assez restrictives) mettre en œuvre des techniques de bootstrap pour, par exemple, tester si un taux de croissance de la PGF est significativement différent de zéro ou si la déviation d'une économie à la frontière technologique mondiale est significative. Simar et Wilson [139] et [136] proposent une présentation générale du bootstrap dans ce cadre, Simar et Wilson [135] proposent une présentation plus spécifique aux indices de croissance de la PGF. L'idée générale du bootstrap est de considérer la suite de frontière estimée sur les données, puis de simuler des échantillons à partir de ces frontières en tirant des niveaux d'efficience dans la distribution empirique de l'efficience. Pour chaque échantillon ainsi formé on ré-estime les frontières et on re-construit l'indice de croissance de la PGF. En répétant on finit par obtenir la distribution de cet indice. Pour que cette technique de bootstrap puisse nous donner une bonne approximation de la distribution de l'estimateur étudié, il faut que des hypothèses très restrictives soient vérifiées : (i) la distribution des *inputs* est indépendante de la distribution de l'efficience, (ii) les paniers d'*inputs* et l'efficience sont des variables aléatoires indépendantes dans la dimension individuelle, (iii) l'efficience est identiquement distribuée pour tous les individus (iv) l'efficience à la date t est indépendante de l'efficience à la date $t+1$. Ces hypothèses sont fortes dans le sens où on a toutes les raisons de penser qu'elles sont violées en pratique. Supposer l'absence de dépendance temporelle contredit l'idée même d'un rattrapage technologique qui énonce l'existence d'une force de rappel des nations vers la frontière. Supposer l'absence de dépendance entre les *inputs* et l'efficience pose aussi problème, car on a l'idée qu'une économie efficace mais faiblement intensive en capital ne conserve pas ses caractéristiques en termes d'efficience si elle décide subitement d'augmenter son niveau de capital par travailleur. Supposer l'homogénéité de la distribution de l'efficience dans la dimension individuelle ne va pas de soi. On peut très bien imaginer qu'en moyenne les économies pauvres soient plus éloignées de la frontière que les économies riches. Enfin, on peut imaginer sans plus de difficulté une dépendance spatiale des observations, en contradiction avec le point (ii). Même si, en relâchant ces hypothèses, on peut toujours montrer la convergence de l'estimateur DEA vers les vraies frontières, la mise en œuvre d'un test (ou, par exemple, la construction d'intervalles de confiance pour les densités estimées dans les graphiques présentés plus haut) doit tenir compte des éventuelles déviations à cet ensemble d'hypothèses. A l'exception du point (ii)[28] Simar et Wilson proposent des amendements à la technique du bootstrap pour tenir compte d'une éventuelle violation de ces hypothèses. Par exemple, pour corrigé de la dépendance temporelle, dans un cas à deux périodes t_1 et t_2, ils proposent de simuler des échantillons en tirant dans la distribution empirique bivariée de l'efficience en t_1 et t_2. Mais ces problèmes ne sont jamais envisagés simultanément. Une application, non reportée, en prenant en compte la dépendance temporelle dans le cas de l'estimateur sta-

[28]Qui à ma connaissance n'est jamais traité.

tique donne des intervalles de confiance très larges pour le taux de croissance de la PGF. Généralement ce taux de croissance n'est pas significatif, et il n'est pas significativement différent de celui que nous obtiendrions en posant que la technologie est Cobb-Douglas et que la part du capital est 0,4. La non significativité des taux de croissance de la PGF tendrait à donner plus de poids à nos conclusions mais il semble difficile d'accorder beaucoup de crédit à ce genre d'exercice (étant donné le nombre d'hypothèses fortes implicitement admises). Le traitement simultané de l'hétérogénéité et de la dépendance temporelle semble difficile à mettre en œuvre, encore plus dans le cas de l'estimateur dynamique de la suite des frontières de production.

Une autre limite du travail présenté dans cette section est la robustesse aux erreurs de mesure (dont le processus générateur des données suppose l'absence) et à la présence d'observations aberrantes. Kneip et Simar [82] suggèrent une procédure en deux étapes pour traiter les erreurs de mesure. Il s'agit dans une première étape d'obtenir un prédicteur non linéaire (à l'aide d'une régression non paramétrique) du produit d'une nation en fonction des facteurs de production, puis de mettre en œuvre l'estimateur DEA en utilisant les *inputs* observés et la prédiction du produit. La première étape permet de retirer le bruit lié aux erreurs de mesure sur le produit. Les deux auteurs ne donnent pas les propriétés asymptotiques de cette procédure d'estimation en deux étapes dans le cas de l'estimateur DEA (mais cela devrait être possible). Un autre problème est qu'ils filtrent les données en estimant le modèle $y_{i,t} = h_i(\mathbf{x}_{i,t}) + \epsilon_{i,t}$ dans la dimension temporelle pour chaque individu. Cela revient à supposer l'absence de progrès technologique, puisque la fonction $h_i(\bullet)$ invariante dans le temps est implicitement liée à la technologie. Dans notre cas se pose aussi le problème de la non stationnarité des variables utilisées. Ainsi il serait peut être plus raisonnable de construire un prédicteur du produit en estimant le modèle dans la dimension individuelle, $y_{i,t} = h_t(\mathbf{x}_{i,t})+\epsilon_{i,t}$. Ceci reste à voir. Cependant, étant données les dimensions limitées des échantillons considérés dans la littérature sur la croissance et l'aspect doublement non paramétrique de cette procédure en deux étapes, il ne serait pas raisonnable de placer beaucoup d'espoir dans une telle approche. Une alternative à l'approche DEA, si on veut tenir compte des erreurs de mesure, est de considérer l'approche dite des *stochastic frontiers*. Cette ap-

proche en panel est paramétrique, le bruit se décompose en une perturbation de moyenne nulle, qui peut capter les erreurs de mesure (sur le produit), et une erreur négative qui correspond au terme d'efficience. La distribution de ces erreurs est généralement spécifiée. Le problème, une nouvelle fois, est le traitement dans ce cadre de la non stationnarité et de l'éventuelle cointégration caractérisant les données sur lesquelles nous travaillons.
Simar [133] propose une procédure pour détecter les observations aberrantes dans l'estimation DEA d'une frontière, mais elle n'est pas automatique et il paraît difficile de l'utiliser si on veut estimer une suite de frontières.

On peut considérer des limites encore plus fondamentales. On a supposé l'existence d'une frontière de production concave unique. L'estimateur dynamique proposé dans cette section n'aurait guère de sens si on relâchait cette hypothèse. Si les ensembles de production ne sont pas convexes, comment admettre que les nations puissent considérer des combinaisons convexes de différents ensembles de production ? Si les économies ne peuvent pas mélanger plusieurs technologies contemporaines à un instant donné pour former la frontière de production mondiale, il paraît difficilement justifiable de supposer qu'elles puissent le faire dans le temps. Dans la section suivante on propose un test de l'hypothèse de convexité de l'ensemble de production.
Finalement il convient de noter que ce genre d'exercice de *growth accounting* ne peut pas évaluer le rôle de la technologie de façon totalement satisfaisante. En effet, on sait que la dynamique du capital n'est pas indépendante de celle de la technologie, voir le graphique [3.2] et le chapitre [1]. Un exercice de comptabilité de la croissance ne mesure que les contributions directes de l'accumulation du capital et de la PGF à la croissance du produit par travailleur. L'effet induit par la dynamique de productivité sur la dynamique du capital n'est pas contrôlé en considérant la dynamique d'accumulation du capital comme exogène[29]. Ainsi, on sur-évalue le rôle de l'accumulation du capital[30]. Il n'existe pas d'approche empirique non-paramétrique réellement

[29]Cette approximation est importante dans la mesure où même un modèle de croissance néo-classique augmenté d'une dynamique de rattrapage technologique exogène, comme dans le chapitre [1], prédit un effet positif du niveau de l'efficience sur la dynamique du capital.

[30]Notons cependant que le sens du biais n'est pas clair si on considère un modèle plus général que dans le chapitre [1], c'est à dire un modèle où les dynamiques de progrès technique et d'efficience sont endogènes. Hulten [75] qui le premier remarqua la limite

satisfaisante pour traiter ce point. On pourrait considérer une efficience non radiale, c'est à dire des distances à la frontière technologique qui ne sont pas orthogonales à l'ensemble engendré par le vecteur des *inputs*. En adoptant une mesure radiale on mesure la distance à la frontière pour un ratio capital–travail donné. On pourrait tout aussi bien considérer une efficience Harrod–neutre. Il s'agit de la spécification retenue par Hall et Jones [69], cette idée apparaît déjà dans Hulten [75]. Ce dernier propose de considérer une mesure harrodienne de la croissance de la PGF. En laissant constant le ratio capital–produit, la mesure harrodienne de la croissance de la PGF (contrairement à la mesure proposée par Solow [141]) attribue une part de la croissance du facteur accumulable à la croissance de la PGF. Mais cette mesure n'est exacte que si les chocs sur la frontière sont effectivement Harrod–neutre. De plus, si l'intérêt de la suggestion de Hulten est transparent dans le cadre de l'approche de Solow, le rapport à l'approche développé dans ce chapitre est moins clair. L'approche proposée dans ce chapitre distingue la croissance de la frontière et la croissance des distances à la frontière. La déformation de la frontière au cours du temps n'est pas contrainte mais la croissance de l'efficience est hicksienne. Il paraît difficile de contraindre la déformation de la frontière de production au cours du temps, mais nous pourrions sans problème considérer une mesure harrodienne de l'efficience dans le cadre de l'approche empruntée dans ce chapitre[31]. Une approche plus convaincante consisterait en l'estimation complète d'un modèle de croissance. Il s'agit de la méthode la plus naturelle pour prendre en compte l'endogénéité de la dynamique du capital physique, puisque seul un modèle peut clairement spécifier la nature du lien entre capital et productivité. Une telle approche souffrirait des défauts d'une approche totalement paramétrique (*ie* quel modèle doit-on choisir ?).

des exercices de comptabilité de la croissance, considère un modèle où le choc de progrès technique est exogène.

[31] Voir Färe et Lovell [57] sur les mesures non radiales de l'efficience.

4 De la convexité de l'ensemble de production mondial

Dans cette section, on revient sur une hypothèse admise dans les sections précédentes et généralement retenue dans la littérature sur la croissance empirique : la concavité de la fonction de production mondiale. Le modèle de croissance néo-classique repose sur cette hypothèse, elle est nécessaire pour montrer l'unicité de l'état stationnaire. Dans une perspective internationale, cette hypothèse est essentielle pour obtenir le résultat de convergence. Ainsi Barro et Sala-i-Martin et leurs successeurs supposent l'unicité et la concavité de la frontière de production mondiale pour donner une interprétation structurelle aux régressions en coupes. Dans les exercices de comptabilité de la croissance, on pose aussi cette hypothèse (voir Easterly et Levine [48], Klenow et Rodriguez-Clare [80], Hall et Jones [69],...).
Les résultats obtenus dans la section et les chapitres précédents pourraient s'expliquer par une violation de l'hypothèse de concavité et d'unicité de la fonction de production mondiale. Nous trouvons dans la section précédente que la dynamique d'accumulation du capital est non linéaire et qu'elle semble cohérente avec deux régimes de convergence. Dans l'introduction générale on a montré, en considérant un panel de nations plus grand, que la dynamique du produit par tête est elle aussi non linéaire. Encore une fois on observe deux régimes de convergence et un régime de divergence. La littérature empirique sur la convergence fournit par ailleurs de multiples exemples.

Par exemple, on peut citer Durlauf et Johnson [45]. Ces auteurs tentent d'aménager l'*approche classique* de la convergence pour tenir compte de l'éventuelle existence de régimes multiples. Ils mettent en œuvre des *regressions tree* basées sur le modèle proposé par Mankiw, Romer et Weil [95] afin de déterminer, de façon endogène, le nombre de clubs de convergence (ou de divergence) et leurs compositions. Leurs résultats sont assez décevants. Ils obtiennent quatre groupes de nations relativement homogènes, qui réservent cependant quelques surprises (comme la position du Japon qui se retrouve dans le même groupe que Madagascar). Leurs conclusions sur la convergence ou la divergence au sein de chaque groupe sont assez étonnantes. Ils ne parviennent pas à conclure à la convergence pour le groupe des économies les plus riches, ce qui, au regard du reste de la littérature, semble difficile à admettre. La division du monde proposée par ces deux auteurs est linéaire (même s'ils considèrent deux variables discriminantes le revenu initial et le capital humain). Desdoigts [36] adopte une approche non linéaire et non paramétrique, en considérant 9 variables susceptibles d'être liées à la croissance, pour former les groupes de nations. La généralité de son approche est coûteuse, relativement au travail de Durlauf et Johnson, (*i*) il ne propose pas de référence explicite à un modèle alors que Durlauf et Johnson considèrent la spécification de MRW [95][32] ; (*ii*) il ne propose pas d'estimation des seuils qui définissent les groupes de nations[33] ; (*iii*) les groupes de nations ne sont pas véritablement déterminés de façon endogène puisque c'est l'observateur qui, en dernier ressort, doit décider de la composition des groupes. Néanmoins les groupes proposés par Desdoigts [36] contredisent moins l'intuition que ceux reportés par Durlauf et Johnson. A l'issue de son analyse il obtient deux groupes clairement distincts et relativement homogènes en termes de croissance. Ainsi les déterminants considérés par Desdoigts pour former les groupes de nations semblent bien expliquer l'hétérogénéité des dynamiques de croissance.

Les faits empiriques ne manquent pas pour étayer l'idée d'une dynamique non linéaire de la croissance. Cette idée de non linéarité de la convergence est fondamentalement différente de la notion de convergence conditionnelle. Dans le premier cas, la condition initiale sur le produit par travailleur est déterminante pour définir l'état stationnaire d'une économie, alors que dans le second cas l'hétérogénéité sur les états stationnaires est déterminée par d'autres variables (institutions, ouverture,...). Comme l'ont montré Bernard et Durlauf [20], une régression en coupe (ou même une analyse en panel) ne peut pas discriminer entre ces deux histoires. Même l'analyse non linéaire de la β-convergence proposée dans l'introduction générale, ne peut discriminer entre ces deux histoires. La démarche, initiée par les travaux de Quah ou Desdoigts [35], qui étudie directement l'évolution de la distribution du produit par tête peut offrir une lecture plus riche des données mais ne peut pas

[32]Les groupes obtenus par Durlauf et Johnson peuvent alors s'apparenter à des clubs de convergence ou de divergence, ce qui n'est pas le cas chez Desdoigts.

[33]Notons que les propriétés asymptotiques des seuils estimés par Durlauf et Johnson ne sont pas plus claires. Voir Hansen [71].

plus apporter de réponses à cette question. Que l'on adopte cette dernière démarche ou l'approche non linéaire de l'introduction générale, on peut toujours trouver un conditionnement tel que la multi-modalité de la distribution du produit par travailleur ou l'aspect non linéaire mis en évidence dans l'introduction générale disparaisse. L'approche par les séries temporelles n'apporte pas plus d'information sur ce point. Aucune démarche empirique directement axée sur l'évaluation de l'hypothèse de convergence ne semble donc pouvoir apporter d'éléments pour trancher entre ces deux histoires.

Dans cette section, on abandonne cette direction et on en vient directement au test d'une condition susceptible de faire pencher la balance d'un côté ou de l'autre. En effet, si les données rejettent l'hypothèse d'une fonction de production mondiale concave, alors on sait que cela peut entraîner une multiplicité d'équilibres et donc faire pencher la balance du côté de la croissance non linéaire et des clubs de convergence. L'idée qui consiste à se concentrer sur l'estimation directe[34] de la fonction de production apparaît déjà dans Duffy et Papageorgiou [43]. Ces deux auteurs proposent d'estimer une fonction de production agrégée CES dans une dimension internationale[35]. Avant de présenter leurs estimations, ils rappellent quelques modèles stylisés de croissance pour souligner que ce genre de modèle peut engendrer une dynamique de long terme endogène ou des équilibres multiples selon la valeur des paramètres de la fonction CES (essentiellement l'élasticité de substitution entre les différents *inputs*). Le travail empirique de ces deux auteurs est doublement insatisfaisant. En premier lieu les données utilisées par ces deux auteurs ne sont pas comparables (il n'y a pas de correction pour la parité des pouvoir d'achat, comme dans Summers et Heston [146]). En second lieu, les auteurs ne s'interrogent pas une seule fois sur les propriétés des séries utilisées. En particulier il n'y a aucune discussion sur l'ordre d'intégration et l'existence de relations de cointégration entre les différents agrégats considérés. Ainsi leurs résultats ne sont pas fiables (pas plus le niveau des estimations que les intervalles de confiance et tests). Duffy et Papageorgiou [43] concluent

[34]Notons que l'intérêt de l'estimation d'une vitesse de convergence réside souvent dans l'information qu'elle peut nous donner sur les paramètres de la fonction de production (*ie* discussion sur le paramètre α de la Cobb-Douglas).

[35]Ils considèrent un panel de 82 nations observées annuellement entre 1960 et 1987. Ces données sont extraites de Nehru et Dhareshwar [100] et Nehru, Swanson et Dubey [101].

en affirmant que les empiristes de la croissance (mais aussi les théoriciens) devraient considérer d'autres spécifications que la fonction de production Cobb-Douglas. Mais pourquoi s'arrêter à une spécification CES ? Ces deux auteurs ne prouvent pas que cette représentation soit satisfaisante. On peut même voir, d'une certaine façon, qu'ils prouvent le contraire. En effet, en découpant leur échantillon de 82 nations en quatre sous-échantillons selon le niveau de développement des économies considérées, ils notent que l'élasticité de substitution entre les facteurs de production dépend du niveau de développement. Dès lors, une fonction de production agrégée imposant la constance de l'élasticité de substitution n'est plus satisfaisante. Dans cette section on adopte une approche plus générale car non paramétrique. Cet accroissement de généralité nous empêche de directement discuter la valeur d'une élasticité de substitution[36]. Néanmoins, on sait que le rejet de la convexité de l'ensemble de production est une propriété susceptible d'impliquer l'existence de régimes multiples de croissance.

Notons toutefois que la non concavité de la frontière de production mondiale peut avoir de multiples explications. Même en adoptant une approche totalement non paramétrique, on ne se protège pas contre une mauvaise spécification de la fonction de production (comme, par exemple, un problème d'agrégation du travail ou du capital). La spécification de la fonction de production mondiale devrait faire l'objet de tests, mais les données manquent pour une aussi large coupe d'économies. Dans la suite on supposera que la non convexité ne vient pas d'un problème de spécification.

La section [4.1] présente un estimateur de la frontière technologique ne reposant pas sur l'hypothèse de convexité de l'ensemble de production. Il s'agit de l'estimateur introduit par Deprins, Simar et Tulkens [34]. La section [4.2] propose un test de l'hypothèse de convexité. Simar et Wilson [137] proposent des statistiques pour sélectionner les spécifications CRS, VRS ou NIRS dans le cas convexe. On étend leur démarche pour tester une spécification convexe contre une spécification non convexe de l'ensemble de production. La section

[36] Il paraît difficile de calculer l'élasticité de substitution à partir de l'estimation de la frontière. En effet, l'estimateur de la frontière est linéaire par morceaux ce qui induit localement une parfaite substitution entre les *inputs*. Par contre on peut calculer la part du capital dans le produit à partir de l'estimation de la frontière.

[4.3] conclut.

4.1 L'estimateur FDH

Il est possible d'estimer une frontière de production sans poser la convexité. Une approche, statistiquement fondée[37], est proposée par Deprins, Simar et Tulkens [34]. Elle permet d'estimer une frontière de production croissante en ses arguments de façon totalement non paramétrique. La méthode d'estimation est proche de l'approche DEA. L'estimateur DEA de la frontière est obtenu en construisant le plus petit ensemble convexe enveloppant les données et vérifiant que l'on peut toujours produire une quantité y avec plus d'*inputs* et produire une quantité inférieure à y avec moins d'*inputs*. Si le couple $(\mathbf{x}, y)$ appartient à l'ensemble de production, alors, pour tout $\lambda_1 > 1$ et $\lambda_2 \in]0, 1]$, $(\mathbf{x}, \lambda_2 y)$ et $(\lambda_1 \mathbf{x}, y)$ appartiennent aussi à l'ensemble de production convexe. On appelle *free disposal* cette propriété de l'ensemble de production. Deprins, Simar et Tulkens [34] proposent un estimateur pour le plus petit ensemble enveloppant les observations vérifiant cette dernière propriété, l'estimateur Free Disposal Hull (FDH).

Définition 4.1. *Soient $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$ les plans de production des nations observées. L'estimateur FDH de l'ensemble de production est défini de la façon suivante,*

$$\widehat{\Psi}_{FDH} = \{(\mathbf{x}, y) \mid \exists\ n \textit{ tel que } \mathbf{x} \geq \mathbf{x}_n \textit{ et } y \leq y_n\} \tag{3.6}$$

La figure [3.10] représente l'estimateur FDH dans le cas avec un facteur de production et un produit. L'ensemble de production FDH appartient à l'ensemble DEA. La distance à la frontière FDH est obtenue plus simplement que dans le cas de la frontière DEA car il n'est pas nécessaire de résoudre un problème de programmation linéaire. Notons $\mathcal{D}_n^{FDH} \in]0, 1]$ la distance de la nation n à la frontière FDH. Si $\mathcal{D}_n^{FDH} = 1$ alors l'économie n est sur la frontière technologique mondiale, sinon cette économie n'est pas efficace. Comme dans le cas de la frontière DEA, cette distance nous apprend par combien nous devrions multiplier le produit d'une économie pour que celle-ci

[37]Voir Park, Simar et Weiner [107].

soit efficace. On obtient un estimateur de la distance à la frontière FDH en considérant le problème suivant,

$$\left(\widehat{\mathcal{D}}_n^{FDH}\right)^{-1} = \max\left\{\theta|(\mathbf{x},\theta y) \in \widehat{\Psi}_{FDH}\right\} \tag{3.7}$$

En pratique, on doit résoudre le programme suivant,

$$\left(\widehat{\mathcal{D}}_n^{FDH}\right)^{-1} = \max_{i\in D_n}\frac{y_i}{y_n} \tag{3.8}$$

avec $D_n = \{i|x_i \leq x_n\}$. Pour faciliter la comparaison avec l'estimateur DEA de la distance à la frontière, on peut réécrire cet estimateur sous la forme d'un problème de programmation linéaire,

$$\begin{aligned}
\left(\widehat{\mathcal{D}}^{FDH}(\mathbf{x}_n, y_n)\right)^{-1} &= \max_{\theta,\{\lambda_i\}_{i=1}^N} \theta \\
\text{Sous les contraintes, } \theta y_n &\leq \sum_{i=1}^N \lambda_i y_i \\
x_{n,m} &\geq \sum_{i=1}^N \lambda_i x_{i,m} \text{ pour } m = 1,\ldots,M \\
\sum_{i=1}^N \lambda_i &= 1 \text{ et } \lambda_i \in \{0,1\} \text{ pour } i = 1,\ldots,N
\end{aligned}$$

Park, Simar et Weiner [107] ont montré sous quelles conditions cet estimateur converge vers la distance à la vraie frontière. Pour obtenir ce résultat il faut, comme pour l'estimateur DEA, définir proprement le processus générateur des données (c'est à dire préciser les propriétés de la frontière technologique et de l'efficience, en particulier, il est toujours important de supposer qu'il existe une probabilité strictement positive d'observer des réalisations arbitrairement proches de la vraie frontière). Ils montrent que la vitesse de convergence est généralement inférieure à celle de l'estimateur DEA, ce résultat est naturel puisque l'on impose moins de structure aux données (on ne suppose plus la convexité de l'ensemble de production). Ils parviennent dans ce cas à dériver la distribution asymptotique de la distance à la frontière. L'estimateur, correctement normalisé par la taille de l'échantillon, tend vers une loi de Weibull qui dépend de paramètres de nuisances inobservés (la pente de la vraie frontière et la probabilité d'observer des réalisations du produit dans un voisinage de la frontière). Ces résultats permettent, par exemple, de construire des intervalles de confiance asymptotiques pour les distances à la frontière et de tester si une nation est inefficace. Néanmoins les paramètres de nuisances ne sont pas simples à estimer, et l'approximation asymptotique peut se révéler insatisfaisante (encore plus dans un cadre non paramétrique). Dans la section suivante nous utiliserons une technique de bootstrap pour mettre en œuvre des tests.

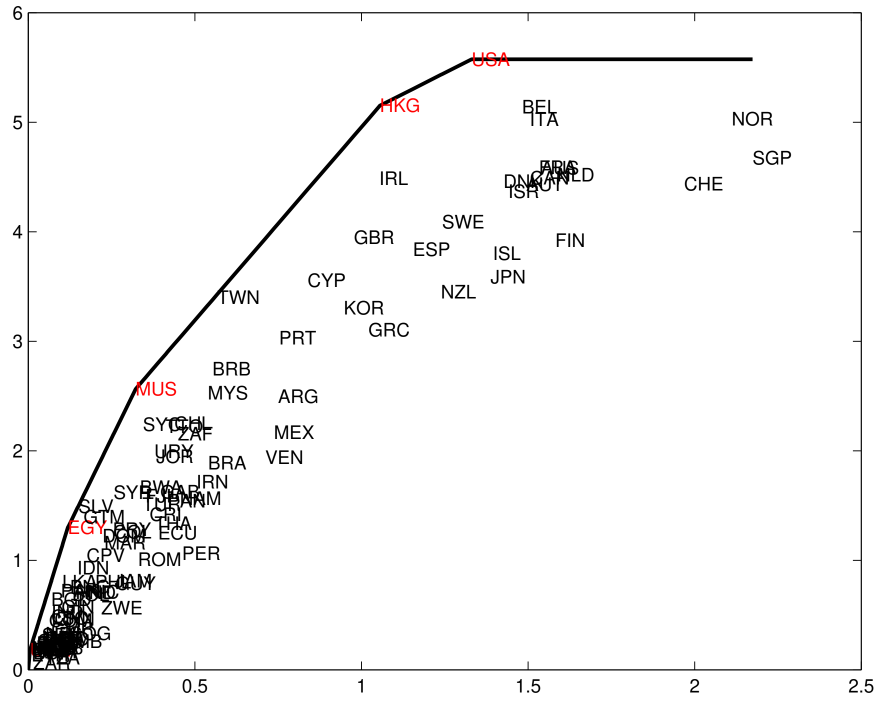

Fig. 3.5 – **Estimation de la frontière NIRS.**

On applique cet estimateur à un échantillon de 107 économies[38] observées en 1995. Les graphiques [3.5] et [3.6], représentent les frontières de production estimées FDH et NIRS dans le plan capital par travailleur – produit par travailleur[39]. Les tableaux [3.8, 3.9 et 3.10 en pages 245, 246 et 247] donnent

[38] L'échantillon est décrit dans la section [5.1, en page 222].

[39] On suppose que la fonction de production est globalement à rendements constants. Cette hypothèse implique que la technologie intensive (dans le plan capital par travailleur

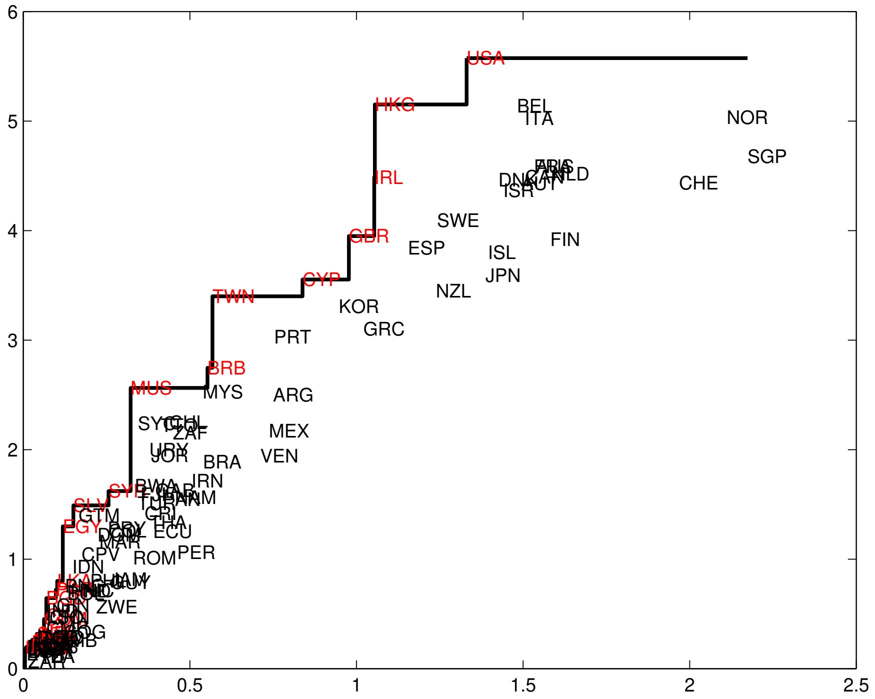

Fig. 3.6 – **Estimation de la frontière FDH.**

nos estimations des distances à la frontière dans les cas NIRS et FDH. Par construction les distances à la frontières FDH sont plus importantes que les distances à la frontière DEA. On observe 5 nations sur la frontière NIRS contre 24 nations sur la frontière FDH. Les figures [3.19, en page 250] et [3.20, en page 251] décrivent des estimateurs de la densité de la distribution des distances aux frontières NIRS et FDH. Clairement, la masse de probabilité est plus à droite dans le cas de la frontière FDH. On remarque aussi que la densité correspondant au cas FDH, contrairement à celle correspondant au cas NIRS, est monotone croissante. Dans ce cas la probabilité d'observer une efficience dans un voisinage de $\theta \in]0,1]$ est d'autant plus grande que θ est proche de l'unité. Les niveaux moyens de l'efficience dans les cas NIRS et FDH sont respectivement 0,63 et 0,77.

– produit par travailleur) est à rendements décroissants. Ainsi, on a utilisé l'estimateur NIRS pour obtenir le graphique [3.5]. L'estimateur FDH est lui aussi appliqué sur des données intensives.

Une question intéressante est de savoir quelles sont les contributions respectives de l'efficience et du capital par travailleur dans le caractère multimodal du produit par travailleur en 1995 (voir la figure [3.21, en page 252]). On considère pour cela les distances estimées à la frontière FDH, car cet estimateur converge vers la vraie frontière de production même si le vrai ensemble de production est caractérisé par une propriété de non convexité. La question est de savoir si l'hétérogénéité des niveaux observés du produit par travailleur est plus liée à une hétérogénéité sur les dotations en capital par travailleur ou à une hétérogénéité sur la propension des nations à exploiter la frontière technologique mondiale. On n'apporte pas ici de réponse quantitative à cette question. On compare simplement la distribution du produit par travailleur avec la distribution du produit par travailleur dans le cas où toutes les nations peuvent produire efficacement. Pour cela, il suffit de projeter les observations du produit par travailleur sur la frontière FDH puis d'estimer la distribution du produit par travailleur efficace. La figure [3.22, en page 253] compare les estimateurs de ces deux densités. Clairement, l'hétérogénéité sur les dotations en capital physique explique bien les modes extrêmes de la distribution, mais le centre de la distribution semble plus lié à l'incapacité des nations à exploiter efficacement la frontière technologique.
Notons néanmoins que ce genre d'exercice suppose, comme dans la section précédente, l'exogénéité du capital par travailleur. On en vient maintenant au test de l'hypothèse de convexité de l'ensemble de production mondiale.

4.2 Tester la convexité

Par construction, l'ensemble de production FDH appartient à l'ensemble de production NIRS ; ainsi, le ratio $\mathcal{D}_n^{NIRS}/\mathcal{D}_n^{FDH}$ est nécessairement inférieur ou égal à un pour tout $n = 1, \dots, N$. Les tableaux [3.8, 3.9 et 3.10], qui reportent les estimations des distances aux frontières NIRS et FDH, suggèrent que ce ratio est très souvent largement inférieur à l'unité. En moyenne ce rapport est égal à 0,82. Pour conclure sur la nature convexe ou non de l'ensemble de production, on doit tester si ce ratio moyen est significativement différent de l'unité. On considérera donc la statistique suivante pour tester la convexité de l'ensemble de production[40],

$$S = N^{-1} \sum_{n=1}^{N} \frac{\mathcal{D}_n^{NIRS}}{\mathcal{D}_n^{FDH}}$$

Pour tester $S = 1$ (convexité de l'ensemble de production) contre $S < 1$ (non convexité de l'ensemble de production), il suffit de simuler la distribution de cette statistique sous l'hypothèse nulle de convexité et de situer la statistique estimée $\widehat{S} = 0,82$ par rapport à cette distribution.

On obtient la distribution sous l'hypothèse nulle en utilisant une technique de bootstrap. A partir des couples $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$ observés, on peut estimer la frontière sous l'hypothèse nulle (NIRS) et construire les couples $\{(\mathbf{x}_n, \widehat{y}_{n,0})\}_{n=1}^N$, avec $\widehat{y}_{n,0} = y_n / \widehat{\mathcal{D}}_n^{NIRS}$, qui décrivent la situation des nations projetées sur la frontière technologique sous l'hypothèse nulle. Une procédure de bootstrap pour obtenir la distribution de $\widehat{S}$ sous l'hypothèse nulle et tester $S = 1$ pourrait alors se décomposer en quelques étapes,

1. Simuler un échantillon sous l'hypothèse nulle,

$$\{(\mathbf{x}_n^{(b)}, y_n^{(b)})\}_{n=1}^N = \{(\mathbf{x}_n, D_n \widehat{y}_{n,0})\}_{n=1}^N$$

où D_n est tiré dans la distribution empirique des distances à la frontière NIRS (avec remplacement et en attribuant un poids $1/N$ à chaque distance estimée).

2. Estimer les distances NIRS et FDH pour cet échantillon, $\{\widehat{\mathcal{D}}_n^{NIRS,(b)}\}_{n=1}^N$ et $\{\widehat{\mathcal{D}}_n^{FDH,(b)}\}_{n=1}^N$
3. Calculer la statistique de test $\widehat{S}^{(b)}$
4. Revenir au point 1. autant de fois que nécessaire.
5. Classer les statistiques simulées sous l'hypothèse nulle par ordre croissant. A partir de cette distribution, on peut évaluer la probabilité, sous l'hypothèse nulle, pour que S soit inférieur à $\widehat{S}$. Si cette probabilité est inférieure à, par exemple, 5% alors on rejette l'hypothèse nulle de

[40] Simar et Wilson [137] considèrent d'autres statistiques. En particulier ils considèrent des estimateurs plus robustes que la moyenne, comme la médiane par exemple. On obtient les mêmes résultats en considérant ces statistiques alternatives.

convexité de l'ensemble de production contre la non convexité de l'ensemble de production au seuil de 5%.

Cette procédure de bootstrap serait relativement simple à mettre en œuvre. Malheureusement elle n'est pas convergente. Cela tient à la nature du processus générateur des données considéré. Le problème est que la variable aléatoire sur laquelle nous basons l'exercice d'échantillonnage, *ie* l'efficience, est définie sur un support borné. Par hypothèse l'efficience est une variable aléatoire continue dans $]0,1]$. En particulier le processus générateur des données suppose que la probabilité d'observer le produit d'une nation sur la frontière technologique est nulle (puisqu'il s'agit d'une variable aléatoire continue), mais par construction notre procédure d'estimation conclut à l'efficacité de certaines nations (comme les États Unis). Ceci explique pourquoi il est impossible de reproduire la distribution des distances à la frontière sous l'hypothèse nulle en échantillonnant directement[41] à partir des distances estimées à la frontière. Même dans le cas le plus favorable où la procédure d'estimation conclurait à la présence d'une unique économie sur la frontière, un tel exercice d'échantillonnage conduirait à associer une probabilité strictement positive à l'occurrence de l'événement "une nation atteint la frontière technologique". En effet, la probabilité que l'efficience échantillonnée ainsi soit égale à un est donnée par

$$\mathbb{P}\left(D_n = 1 | \{(\mathbf{x}_n, y_n)\}_{n=1}^N\right) = 1 - \left(1 - \frac{1}{N}\right)^N \underset{N\to\infty}{\longrightarrow} 1 - e^{-1} = 0,6321 > 0$$

Même si nous décidions d'amender le processus générateur des données pour autoriser une discontinuité de la densité de l'efficience, cette procédure ne serait pas plus satisfaisante car elle conduirait invariablement à associer une probabilité égale au moins à 0,6321 asymptotiquement à l'événement "une nation atteint la frontière technologique" lorsque la taille de l'échantillon tend vers l'infini.

Simar et Wilson [137][42] proposent d'échantillonner en mettant en œuvre

[41] c'est à dire en tirant avec remplacement et équiprobabilité dans la distribution empirique de l'efficience.

[42] Cette idée et les remarques précédentes sur la convergence de la procédure de bootstrap apparaissent déjà dans Simar et Wilson [134].

la procédure de *smooth bootstrap* suggérée par Silverman [132, pages 142-147]. Il s'agit d'estimer la densité de l'efficience à l'aide d'un estimateur à noyau et de former les échantillons sous la nulle en tirant l'efficience dans sa distribution estimée. On rencontre alors deux difficultés toujours liées à la nature du processus générateur des données. La variable aléatoire d'efficience est distribuée sur un support borné et l'estimateur à noyau n'est pas adapté dans ce cas[43]. En particulier, la probabilité d'observer des réalisations proches de un est importante. Silverman montre que l'estimateur à noyau est alors biaisé (surtout si on utilise une fenêtre avec un support non borné comme la fenêtre gaussienne). Afin d'obtenir une estimation non biaisée de la densité, Silverman [132, pages 29-32] propose de refléter les observations autour des bornes. Le deuxième problème est lié au choix du paramètre de lissage. On choisit ce paramètre de façon à minimiser l'intégrale de l'erreur quadratique moyenne (MISE, pour Mean Integrated Squared Error) de l'estimateur en utilisant une procédure de *cross-validation*. Silverman [132, pages 51-52] montre que cet estimateur du paramètre de lissage peut tendre vers zéro lorsque les observations prennent éventuellement des valeurs discrètes. Or, par construction de l'estimateur de la frontière, il peut arriver que les réalisations estimées de l'efficience soient plusieurs fois égales à l'unité. Le choix du paramètre de lissage par une procédure de *cross-validation* n'est donc pas adapté à notre cas. Silverman propose de minimiser le critère sur un intervalle de valeurs possibles du paramètre de lissage défini autour de celui que nous obtiendrions avec une *rule-of-thumb*[44]. Simar et Wilson proposent de minimiser une moyenne pondérée de l'intégrale de l'erreur quadratique (MWISE, Mean Weighted Integrated Squared Error). Dans notre application nous avons procédé plus simplement en ne considérant pas les nations efficientes pour déterminer le paramètre de lissage. La figure [3.19, en page 250] représente notre estimation de la densité de l'efficience sous l'hypothèse nulle.

Enfin, notons qu'en formant des échantillons sous l'hypothèse nulle en

[43]Un estimateur basé sur des séries orthogonales serait plus adapté dans ce cas. Nous suivons néanmoins la méthode utilisée par Simar et Wilson [137].

[44]Cette règle de choix automatique pour le paramètre de lissage suppose que les données sont des réalisations d'une variable aléatoire gaussienne. Cette règle est assez peu robuste et particulièrement inadaptée à notre cas.

tirant dans la distribution estimée de l'efficience, on suppose implicitement que celle-ci est orthogonale aux *inputs*. Pour ne pas poser cette indépendance il faudrait échantillonner en tirant dans la distribution estimée du triplet $\left(\mathbf{x}_n, \widehat{y}_n^*, \widehat{\mathcal{D}}_n^{NIRS}\right)$. Ceci ne pose aucun problème technique, mais il n'est pas évident que l'estimation de la densité d'une variable aléatoire de dimension trois sur la base d'une centaine d'observations soit assez précise pour que cet exercice puisse nous apporter une réponse plus robuste.

On a mis en œuvre cette procédure de bootstrap, avec 10000 simulations. La figure [3.7] représente une estimation de la distribution de la statistique de test sous l'hypothèse nulle de convexité de l'ensemble de production. Notons que 1 n'appartient pas à cette distribution, qui paraît plutôt centrée sur 0,9. Ce décalage ne doit pas nous étonner. A distance finie, l'estimateur FDH est plus biaisé que l'estimateur DEA sous l'hypothèse nulle d'une frontière NIRS, puisqu'il n'exploite pas l'information sur la convexité de l'ensemble de production. On remarque aussi que cette distribution est très concentrée. Cette observation n'est guère plus étonnante puisque l'estimateur de la frontière DEA converge généralement plus vite vers la vraie frontière que l'estimateur FDH. Dans le cas général avec q *outputs* et m *inputs*, Kneip, Park et Simar [81] montrent que l'estimateur DEA est un $O_p(N^{-\frac{2}{q+m+1}})$ et Park, Simar et Weiner [107] montrent que l'estimateur FDH est $O_p(N^{-\frac{1}{q+m}})$. Dès lors le ratio des distances estimées est un $O_p(N^{-\frac{q+m-1}{(q+m)(q+m+1)}})$. Ainsi la variance de la distribution du ratio sous l'hypothèse nulle doit tendre vers zéro[45] lorsque la dimension de l'échantillon tend vers l'infini (pour tout $(q, m) \geq (1, 1)$). Dans le cas qui nous intéresse avec $q = m = 1$, on a finalement $\widehat{S} \sim O_p(N^{-\frac{7}{6}})$.
Notre estimateur de la statistique de test ($\widehat{S} = 0.82$) n'appartient pas au support de cette distribution et se situe sur la gauche de cette distribution. L'hypothèse d'une frontière concave de type NIRS est clairement rejetée contre celle d'une frontière non concave quel que soit le seuil considéré. Dans ce test nous avons considéré une frontière convexe de type NIRS sous la nulle. Nous pourrions choisir une spécification CRS sous l'hypothèse nulle, c'est à dire supposer que la fonction de production est linéaire dans le capital par travailleur (cela correspondrait au modèle Ak de croissance endogène). Nous

[45] Il serait plus élégant de construire une statistique de test en posant une normalisation assurant la tendance de la statistique vers une distribution non dégénérée. Bien sûr, cela ne pourrait changer le résultat obtenu.

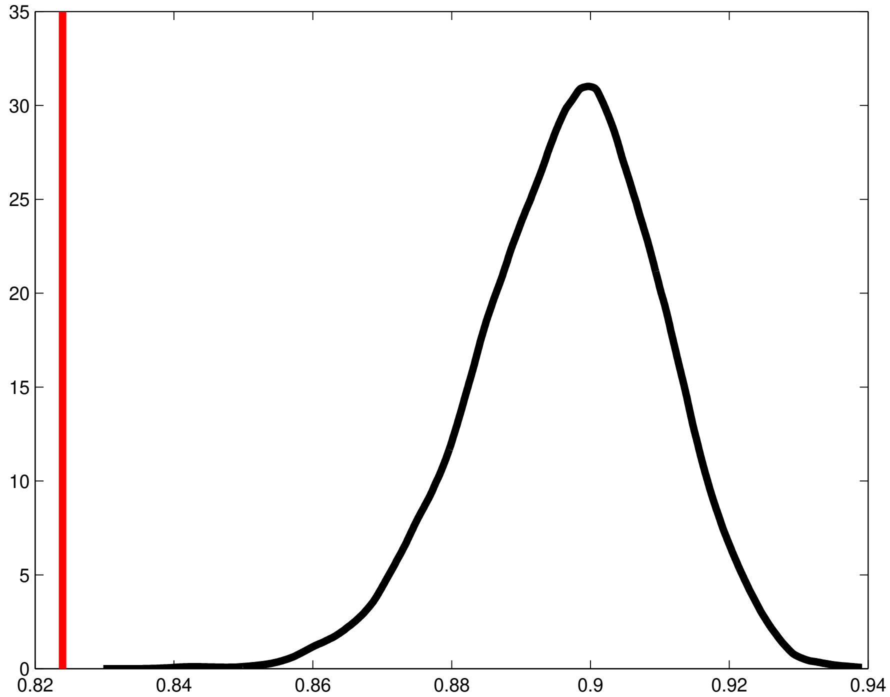

Fig. 3.7 – **Tester la convexité de l'ensemble de production mondial.** La courbe noire correspond à un estimateur non paramétrique de la distribution de la statistique de test sous l'hypothèse nulle (le paramètre de lissage est choisi par une procédure de *cross-validation*). La droite rouge correspond à notre estimation de la statistique de test.

rejetterions de la même façon cette hypothèse nulle. Une spécification VRS sous l'hypothèse nulle donne aussi le même résultat de rejet de la convexité, car les frontières NIRS et VRS sont identiques sur la restriction définie par les combinaisons convexes des *inputs* observés. C'est donc bien la convexité que nous rejetons, et non pas la spécification des rendements d'échelle.

4.3 Conclusion

La conclusion obtenue dans cette section remet en cause le contenu des sections précédentes, mais aussi la mesure indirecte des contributions respectives du capital et de la technologie proposée dans le chapitre [1]. Ce test nous dit que l'hypothèse d'une unique fonction de production mondiale concave est intenable. On ne peut mélanger le charrue utilisée par le laboureur éthiopien et le tracteur climatisé du fermier américain. Ainsi, le processus générateur des données adopté dans la section [2.2] est insatisfaisant. De même, l'approche utilisée pour contraindre la suite des ensembles de production à être croissante repose sur une hypothèse trop forte que les données rejettent. Dans le chapitre [1] nous avons aussi comparé les nations ou régions à une fonction de production commune concave, en suivant la majorité des travaux inscrits dans la littérature consacrée à la croissance et la convergence. Cette hypothèse et donc les résultats dérivés dans ce chapitre sont remis en question.

Il peut paraître assez naïf de supposer que les observations (capital–travail–produit) ne résultent que d'une fonction de production et d'un terme d'efficience. Même s'il est très général, le processus générateur des données considéré ici est beaucoup trop simple. Il est même légitime de se demander si l'estimation d'une fonction de production indépendamment d'un modèle de croissance a un sens. Si on rejette la convexité de l'ensemble de production, faut-il nécessairement en conclure que la fonction de production mondiale est non concave ? Si les données sont générées par un modèle plus complexe que celui supposé par les estimateurs DEA–FDH, la non convexité détectée pourrait provenir d'une autre partie du modèle (comme par exemple des imperfections sur le marché financier). Ce que nous dit le résultat obtenu ici, c'est que le processus générateur des données admet une non convexité, l'identification de l'origine de cette non convexité est une question plus difficile qui paraît difficile d'examiner sans considérer un modèle économique autour de la fonction de production.

5 Annexes

5.1 Les données

les données utilisées dans ce chapitre sont essentiellement extraites des *Penn World Tables* 5.6 et d'une version préparatoire des *Penn World Tables* 6 par Summers et Heston [146].

Pour la section [3] on a utilisé PWT56, cette base offre des données annuelles (comparables dans les dimensions temporelles et spatiales) de 1965 à 1990 sur le capital physique, la force de travail, et le produit réel pour 59 économies. On a utilisé les variables POP (*population*), RGDPCH (*Real Gross Domestic Product per Capita, international Prices*), RGDPW (*Real Gross Domestic Product per Worker, international Prices*) et KAPW (*Physical Capital Stock per Worker, international Prices*). Summers and Heston obtiennent le stock de capital physique en mettant en œuvre des exercices d'inventaire perpétuel sur des données désagrégées d'investissement (*ie*, équipement et machines, structures,...). Le taux de dépréciation est spécifique à chaque type d'investissement, mais est invariant dans la coupe des économies et dans le temps. Les Tableaux [3.1, en page 233] et [3.2, en page 234] résument les caractéristiques de cet échantillon. Deux économies sont exclues de notre échantillon (l'Iran et le Vénézuela) à cause de l'importance du prix du pétrole sur la dynamique de richesse de ces nations. L'inclusion de ces économies changerait les résultats car ces économies apparaîtraient souvent sur la frontière technologique mondiale, selon l'évolution du prix du pétrole. Par exemple, jusqu'à la fin des années 70, l'Iran (la Perse à l'époque) pousse régulièrement la frontière technologique vers le haut.

Pour la section [4], on a utilisé PWT6. Actuellement cette base de données ne fournit pas d'information sur le stock de capital physique. De ce fait, on a construit le stock de capital physique de l'année 1995 à l'aide d'un exercice d'inventaire perpétuel en utilisant la part de l'investissement dans le produit réel, la variable intitulée KI[46]. On a utilisé les variables POP, RGDPCH, RGDPW et KI. Les économies sont sélectionnées en fonction des données disponibles ; on a ainsi retenu 108 économies pour lesquelles on dis-

[46]Voir http ://webhost.bridgew.edu/baten/pwt6/append6.0.pdf.

pose des données nécessaires sur la période 1960-1995[47]. Le problème pour la construction du stock de capital en 1995 est le choix de la condition initiale en 1960. De nombreux auteurs estiment ce niveau initial en considérant qu'à cette date les économies évoluent (depuis déjà très longtemps) le long d'une trajectoire de long terme dont la pente est commune dans la coupe des économies. Par exemple, Hall et Jones [69] posent que la condition initiale du stock de capital est,

$$K_{60} = \frac{I_{60}}{g+\delta}$$

où g est le taux de croissance commun du progrès technique[48] et δ le taux de dépréciation du capital physique. Cette façon de procéder est assez génante puisque (i) elle revient à supposer que les économies évoluent le long d'une trajectoire de long terme depuis l'origine du monde et (ii) elle revient aussi à supposer que le modèle de croissance néo-classique est une bonne représentation du monde (les économies évoluent le long de trajectoires de long terme parallèles). Easterly et Levine [48] introduisent du bruit en pondérant le taux de croissance g avec un taux de croissance spécifique à chaque économie. Plutôt que de se donner une condition initiale, on a préféré procéder comme Summers et Heston en reconstituant les séries d'investissement jusqu'en 1950 (pour cela on a considéré le taux de croissance moyen, spécifique à chaque nation, de l'investissement entre 1960 et 1970). Le stock de capital physique en 1995 est la somme des flux d'investissement pondérés par les puissances de $(1-\delta)$,

$$K_{1995} = \sum_{t=1950}^{1994} (1-\delta)^{1994-t} I_t$$

On a posé $\delta = 0,05$. Les tableaux [3.8, 3.9 et 3.10 en pages 245-247] donnent le ratio du capital physique au produit en 1995. La figure [3.17, en page 248] représente notre échantillon dans le plan capital par travailleur – produit par travailleur en 1995. A la vue de ce graphique, on a éliminé le Luxembourg

[47] Pour deux économies le Nicaragua et le Tchad on n'observe pas la part de l'investissement dans le produit réel en, respectivement, 1980 et 1974. Plutôt que d'éliminer ces deux économies de l'échantillon on a complété la base de données à l'aide d'une approximation par *spline*.

[48] Hall et Jones considèrent le taux de croissance moyen sur la période 1960-1970 des séries d'investissement.

de l'échantillon dont la position semble aberrante. En laissant cette nation dans l'échantillon, les États Unis n'apparaîtraient plus sur la frontière technologique et des économies comme la Suisse ou la Norvège seraient largement plus inefficaces.

5.2 Codes et sources

Dans la section [3], les estimations de la frontière technologique, des indices de croissance de la PGF et de ses décompositions sont obtenues avec un programme écrit en FORTRAN 77. Le code est compilé avec g77 (un compilateur GNU) et on a utilisé la routine LA01BD extraite de la *HSL Archive* pour résoudre les problèmes de programmation linéaire. Voici un extrait du code utilisé pour estimer les frontières technologiques avec l'estimateur statique et l'estimateur dynamique. Les frontières estimées sont stockées dans les variables TEKF1 (estimateur dynamique) et TEKF2 (estimateur statique).

```
      PROGRAM FRONTIER
      INTEGER N,NO,T,COMPT,I,J,TT,NUMBER,OBS,OBSO,TIME
      CHARACTER*10 DFILE
      PARAMETER(N=59,T=26,OBS = N*T,TT=1,NO = 57)
      PARAMETER(OBSO = NO*T)
      PARAMETER(DFILE = 'E59.data')
      DOUBLE PRECISION POP(OBS),RGDPCH(OBS),RGDPW(OBS),KAPW(OBS)
      DOUBLE PRECISION GDP(OBS),WORKER(OBS),KAP(OBS)
      DOUBLE PRECISION Y(T,NO),L(T,NO),K(T,NO)
      DOUBLE PRECISION DIST(TT,OBS)
      DOUBLE PRECISION DISTN(TT,OBS)
      DOUBLE PRECISION TEKF1(T,NO)
      DOUBLE PRECISION TEKF2(T,NO)
      DOUBLE PRECISION TMP1(TT,OBSO)
      DOUBLE PRECISION TMP2(TT,OBSO)
      DOUBLE PRECISION TMP3(TT,OBSO)
      DOUBLE PRECISION CUTOFF
      PARAMETER(CUTOFF = 0.9999999999999 D 0)
C     Extraction des donn\'{e}es.
      OPEN(UNIT=10,FILE=DFILE,ACCESS='SEQUENTIAL',STATUS='OLD',
     * FORM='FORMATTED')
      DO 100, I = 1,OBS
        READ(UNIT = 10,FMT=*) POP(I),RGDPCH(I), RGDPW(I), KAPW(I)
100   CONTINUE
      CLOSE(10)
      DO 200, I = 1,OBS
        GDP(I) = POP(I) * RGDPCH(I) * 1000
        WORKER(I) = GDP(I) / RGDPW(I)
        KAP(I) = KAPW(I) * WORKER(I)
200   CONTINUE
C     Elimination de l'Iran et du V\'{e}n\'{e}zuela.
```

```
      COMPT = 1
      MINUS = 0
      DO 300, I = 1,N
        IF (I .NE. 29 .AND. I .NE. 26 ) THEN
          DO 400, J = 1,T
            Y(J,I-MINUS) = GDP(COMPT)
            K(J,I-MINUS) = KAP(COMPT)
            L(J,I-MINUS) = WORKER(COMPT)
            COMPT= COMPT + 1
400       CONTINUE
        ELSE IF (I .EQ. 29 .OR. I .EQ. 26) THEN
          DO 410, J = 1,T
            COMPT = COMPT + 1
410       CONTINUE
          MINUS = MINUS + 1
        END IF
300   CONTINUE

C === STEP 1 === Initialisation.
      DO 500, I = 1,OBS0
          TMP1(TT,I) = 0.0 D 0
          TMP2(TT,I) = 0.0 D 0
          TMP3(TT,I) = 0.0 D 0
500   CONTINUE
      DO 505, I = 1,N0
        TMP1(TT,I) = Y(1,I)
        TMP2(TT,I) = K(1,I)
        TMP3(TT,I) = L(1,I)
505   CONTINUE
      NUMBER = N0
      CALL DEA0(OBS0,NUMBER,TMP1,TMP2,TMP3,TMP1,TMP2,TMP3,DIST)
      DO 510, I = 1,NUMBER
        WRITE(*,*) DIST(1,I)
        TEKF1(1,I) = Y(1,I) / DIST(1,I)
        WRITE(*,*) TEKF1(1,I)
        WRITE(*,*) DIST(1,I)
        TEKF2(1,I) = Y(1,I) / DIST(1,I)
        WRITE(*,*) TEKF2(1,I)
510   CONTINUE

C === STEP 2 === Loop !
      DO 600, TIME = 2, T
        IF(TIME .NE. 2) THEN
          I = N0 + 1
10        IF(I .LE. NUMBER) THEN
            IF(I .LT. NUMBER) THEN
              IF(DIST(1,I) .LE. CUTOFF) THEN
                DO 20, J = I,NUMBER-1
                  DIST(1,J) = DIST(1,J+1)
                  TMP1(1,J) = TMP1(1,J+1)
                  TMP2(1,J) = TMP2(1,J+1)
                  TMP3(1,J) = TMP3(1,J+1)
20              CONTINUE
                DIST(1,NUMBER) = 0.0 D 0
```

```
            TMP1(1,NUMBER) = 0.0 D 0
            TMP2(1,NUMBER) = 0.0 D 0
            TMP3(1,NUMBER) = 0.0 D 0
            NUMBER = NUMBER - 1
            GO TO 10
          ELSE IF(DIST(1,I) .GT. CUTOFF) THEN
            I = I + 1
            GO TO 10
          END IF
        ELSE
          IF(DIST(1,I) .LE. CUTOFF) THEN
            DIST(1,I) = 0.0 D 0
            TMP1(1,I) = 0.0 D 0
            TMP2(1,I) = 0.0 D 0
            TMP3(1,I) = 0.0 D 0
            NUMBER = NUMBER - 1
            GO TO 10
          END IF
        END IF
      END IF
    END IF
    DO 630, I = 1,NO
      TMP1(TT,I) = Y(TIME,I)
      TMP2(TT,I) = K(TIME,I)
      TMP3(TT,I) = L(TIME,I)
630 CONTINUE
    J = 0
    DO 640, I = 1,NO
      IF(DIST(1,I) .GT. CUTOFF) THEN
        J = J + 1
        TMP1(TT,NUMBER+J) = Y(TIME-1,I)
        TMP2(TT,NUMBER+J) = K(TIME-1,I)
        TMP3(TT,NUMBER+J) = L(TIME-1,I)
      END IF
640 CONTINUE
    NUMBER = NUMBER + J
    CALL DEA0(OBS0,NUMBER,TMP1,TMP2,TMP3,TMP1,TMP2,TMP3,DIST)
    CALL DEA0(OBS0,NO,TMP1,TMP2,TMP3,TMP1,TMP2,TMP3,DISTN)
    DO 650, I = 1,NO
      TEKF1(TIME,I) = Y(TIME,I) / DIST(1,I)
      TEKF2(TIME,I) = Y(TIME,I) / DISTN(1,I)
650 CONTINUE
600 CONTINUE

C === STEP 2 === Sauvegarde
      OPEN(UNIT=50,FILE='TEKF1.dat',STATUS='NEW')
      WRITE(UNIT=50,FMT = 5) ((TEKF1(J,I),I=1,NO),J=1,T)
      CLOSE(50)
      OPEN(UNIT=60,FILE='TEKF2.dat',STATUS='NEW')
      WRITE(UNIT=60,FMT = 5) ((TEKF2(J,I),I=1,NO),J=1,T)
      CLOSE(60)
5     FORMAT(1X,26(/,1X,57 D12.4))

C === Note that 57 = 59 - 2(VENEZUELA + IRAN)
```

```
      END
C === ****************************************************************
C === ****************************************************************
C === ****************************************************************
      SUBROUTINE DEA0(NN,NUMBER,Y,X1,X2,YY,XX1,XX2,ODF)
      INTEGER NN,NUMBER
      DOUBLE PRECISION Y(1,NN), X1(1,NN), X2(1,NN), ODF(1,NN)
      DOUBLE PRECISION YY(1,NN), XX1(1,NN), XX2(1,NN)
      INTEGER M, L, IA, IPR, IER, I, ECO
      PARAMETER(M = 2, L = 2, IA = M, IPR = 0)
      INTEGER INDEX1,INDEX2
      PARAMETER(INDEX1 = M+1, INDEX2 = INDEX1*(M+3))
      INTEGER IND(INDEX1)
      DOUBLE PRECISION WK(INDEX2), A(IA,NUMBER), B(IA,1), C(NUMBER)
      DOUBLE PRECISION Z(NUMBER), F
      DO 100, ECO = 1,NUMBER
        CALL ONESD(IA,1,B)
        DO 10, I = 1,NUMBER
          C(I)   = -YY(1,I)/Y(1,ECO)
          A(1,I) = XX1(1,I)/X1(1,ECO)
          A(2,I) = XX2(1,I)/X2(1,ECO)
10      CONTINUE
        CALL LA01BD(NUMBER,M,L,A,B,C,Z,F,IA,IPR,IND,WK,IER)
        IF (IER .LE. 0) THEN
          ODF(1,ECO) = -1.0 D 0 / F
        ELSE
          ODF(1,ECO) = SQRT(-2.0 D 0)
        END IF
100   CONTINUE
      END
C === ****************************************************************
      SUBROUTINE ONESD(LIGNE,COLON,TMP)
      INTEGER LIGNE, COLON, I, J
      DOUBLE PRECISION TMP(LIGNE,COLON)
      DO 10, I = 1,COLON
        DO 20, J = 1,LIGNE
          TMP(J,I) = 1.0 D 0
 20     CONTINUE
 10   CONTINUE
      END
C === ****************************************************************
C === ****************************************************************
C === ****************************************************************
```

Dans la section [4], les estimations de la frontière sont obtenues à partir de MATLAB 5.3 et de sa *toolbox* d'optimisation (version 2).

5.3 Preuve de la proposition [2.1]

On reprend la preuve donnée par Banker [11], le seul changement est lié à l'hypothèse $\mathcal{H}7$ qui ne suppose pas l'indépendance de l'efficience et du panier des *inputs*, comme dans Kneip, Park et Simar [81]. On se limite au cas avec un *input* et un *output*. On n'écrit pas les indices liés au temps, puisque l'on considère ici l'estimateur statique. La figure [3.23, en page 254] accompagne la lecture de la preuve.

Rappelons que la fonction de production $\mathcal{F}(\bullet)$ est continue sur la restriction convexe X, sous les hypothèses $\mathcal{H}1$–$\mathcal{H}3$. Ainsi, pour tout x_0 à l'intérieur de l'ensemble X et tout $\Delta > 0$ il existe $\delta > 0$ tel que $\mathcal{F}(x) > \mathcal{F}(x_0) - \Delta$ pour tout $x \in]x_0 - \delta, x_0 + \delta[\subseteq X$.

Pour montrer que l'estimateur DEA, $\widehat{\mathcal{F}}(\bullet)$, est convergent sur la restriction X, on doit montrer que lorsque la taille de l'échantillon tend vers l'infini, la probabilité que la différence entre $\mathcal{F}(x_0)$ et $\widehat{\mathcal{F}}(x_0)$ soit plus grande qu'une constante positive Δ tend vers zéro.

Soit une réalisation (x, y) avec $y = \mathcal{F}(x) - \varepsilon$, où x appartient à la restriction convexe X et ε est une réalisation positive de l'efficience. x et ε sont réalisations de deux variables aléatoires qui ne sont pas nécessairement indépendantes. On considère deux événements $\mathcal{A}_1$ et $\mathcal{A}_2$ définis de la façon suivante,

$$\mathcal{A}_1 = \text{“ } x \in]x_0 - \delta, x_0[\text{ et } y = \mathcal{F}(x) - \varepsilon > \mathcal{F}(x_0) - \Delta \text{ ”}$$

$$\mathcal{A}_2 = \text{“ } x \in]x_0, x_0 + \delta[\text{ et } y = \mathcal{F}(x) - \varepsilon > \mathcal{F}(x_0) - \Delta \text{ ”}$$

On pose $p_1 = \mathbb{P}(\mathcal{A}_1)$, soit de façon équivalente,

$$p_1 = \mathbb{P}\left(\varepsilon < \mathcal{F}(x) - \mathcal{F}(x_0) + \Delta | x \in]x_0 - \delta, x_0[\right) \times \mathbb{P}\left(x \in]x_0 - \delta, x_0[\right)$$

En notant que $\mathcal{F}(x) - \mathcal{F}(x_0) + \Delta > 0$, on a $p_1 > 0$ par l'hypothèse $\mathcal{H}7$. De la même façon on montre que $p_2 = \mathbb{P}(\mathcal{A}_2)$ est une probabilité strictement positive.

Considérons un échantillon de N observations indépendantes et notons $\mathcal{A}_i^N$ l'événement “Aucune observation vérifie $\mathcal{A}_i$”, pour $i = 1$ ou 2. On a directement $\mathbb{P}\left(\mathcal{A}_i^N\right) = (1 - p_i)^N$ par l'hypothèse d'indépendance. Supposons que parmi les N réalisations on ait deux couples notés (x_1, y_1) et (x_2, y_2) tels que $x_1 \in]x_0 - \delta, x_0[$, $x_2 \in]x_0, x_0 + \delta[$ et $y_i \geq \mathcal{F}(x_0) - \Delta$ pour $i = 1$ et 2. Il y a donc au moins deux réalisations vérifiant les événements $\mathcal{A}_1$ et $\mathcal{A}_2$. Par construction l'estimateur DEA $\widehat{\mathcal{F}}(x_0)$ est supérieur ou égale au minimum de y_1 et y_2. Ainsi, nous avons,

$$\begin{array}{rcl} \widehat{\mathcal{F}}(x_0) \geq \min\{y_1, y_2\} & > & \mathcal{F}(x_0) \\ \Leftrightarrow \mathcal{F}(x_0) - \widehat{\mathcal{F}}(x_0) & < & \Delta \\ \Leftrightarrow |\mathcal{F}(x_0) - \widehat{\mathcal{F}}(x_0)| & < & \Delta \end{array}$$

où la dernière équivalence vient du fait que par construction l'estimateur DEA sous-estime nécessairement la vraie frontière. Considérons maintenant la probabilité que la différence entre $\mathcal{F}(x_0)$ et $\widehat{\mathcal{F}}(x_0)$ soit plus grande qu'une constante positive Δ arbitrairement petite.

$$\begin{array}{rcl} \mathbb{P}\left(|\mathcal{F}(x_0) - \widehat{\mathcal{F}}(x_0)| > \Delta\right) & = & \mathbb{P}\left(\mathcal{F}(x_0) - \widehat{\mathcal{F}}(x_0) > \Delta\right) \\ & \leq & \mathbb{P}\left(\mathcal{A}_1^N \cup \mathcal{A}_2^N\right) \\ & \leq & \mathbb{P}\left(\mathcal{A}_1^N\right) + \mathbb{P}\left(\mathcal{A}_2^N\right) - \mathbb{P}\left(\mathcal{A}_1^N | \mathcal{A}_2^N\right) \mathbb{P}\left(\mathcal{A}_2^N\right) \\ & \leq & (1 - p_1)^N + (1 - p_2)^N + o(1) \end{array}$$

Finalement on a bien montré que,

$$\lim_{N \Rightarrow \infty} \mathbb{P}\left(|\mathcal{F}(x_0) - \widehat{\mathcal{F}}(x_0)| > \Delta\right) = 0$$

pout tout $\Delta > 0$. Ainsi on a bien une convergence en probabilité, point par point, de l'estimateur DEA vers la vraie frontière de production.

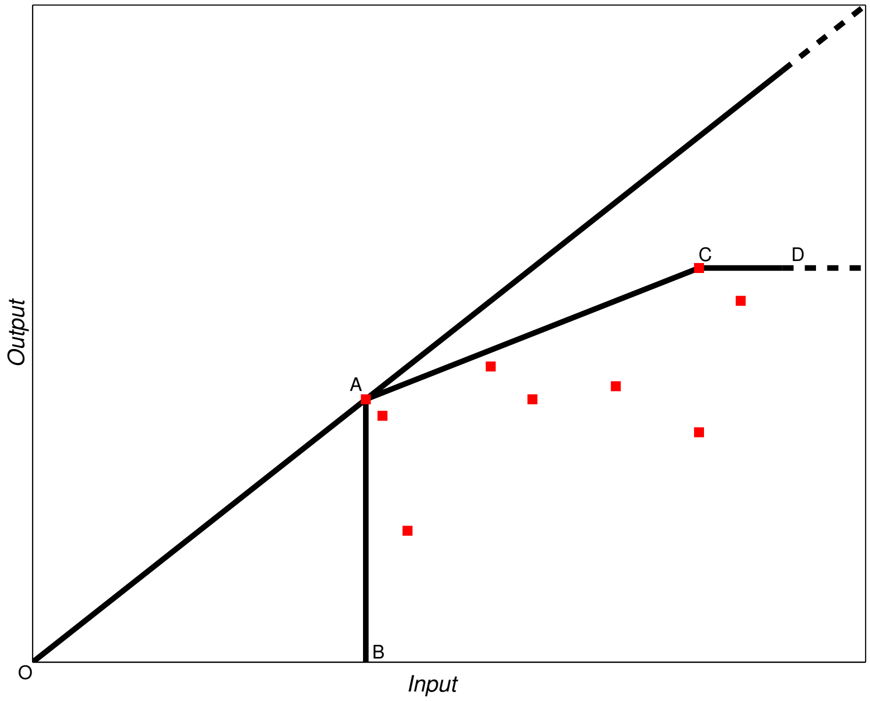

Fig. 3.8 – **Estimateurs de la frontière technologique.** Les courbes noires représentent les frontières de production estimées pour différentes hypothèses sur les rendements d'échelle. Les carrés rouges représentent les observations. La droite $[O, A)$ est la frontière dans le cas CRS, la droite segmentée $[O, A] \cup [A, C] \cup [C, D)$ est la frontière NIRS et la droite segmentée $[B, A] \cup [A, C] \cup [C, D)$ correspond à la frontière VRS.

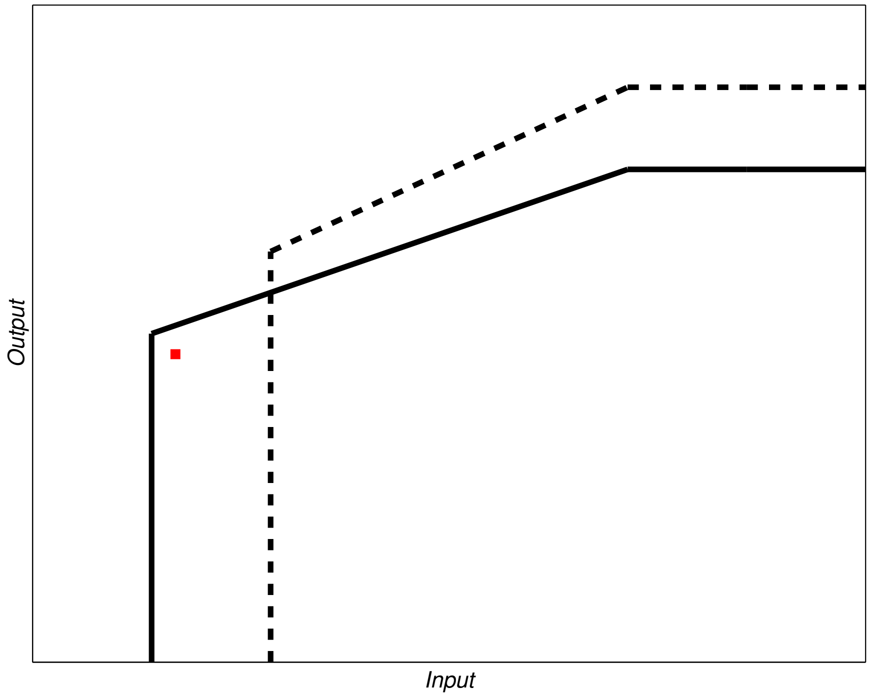

Fig. 3.9 – **Spécification VRS et identification.** La courbe noire en trait plein correspond à la limite de l'ensemble de production de la date t. La courbe noire en tirets définit l'ensemble de production à la date $t+h$, $h > 0$.

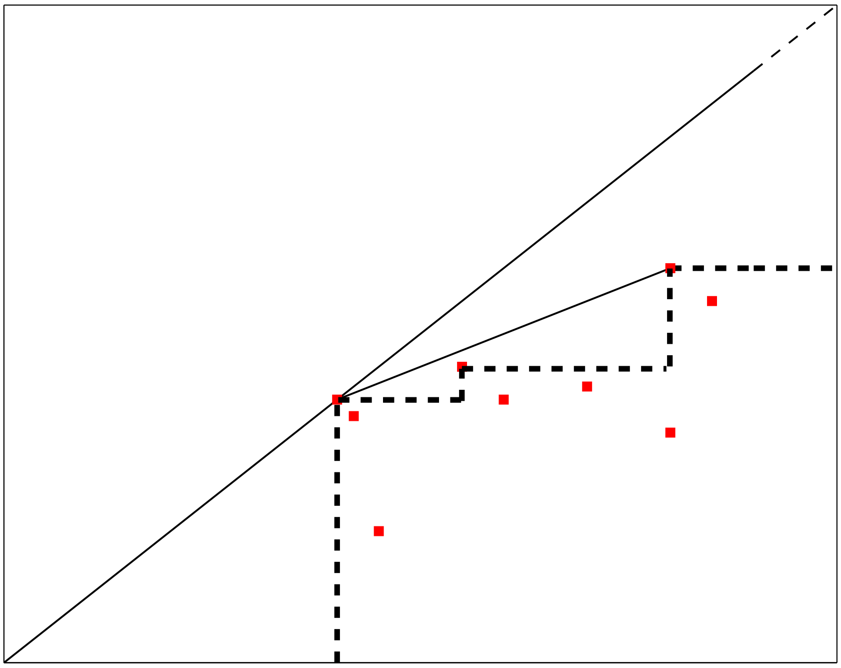

FIG. 3.10 – **Estimateur FDH de la frontière.** On reprend les même observations que la figure [3.8]. La courbe noire en tirets définit l'ensemble de production à la date t, obtenue avec un estimateur FDH.

Tab. 3.1 – Description des données (Niveau)

	Y	L	K/Y
Ivory Coast [CIV]	0.0126	0.0034	0.2161
Kenya [KEN]	0.0131	0.0068	0.6091
Madagascar [MDG]	0.0076	0.0039	0.8963
Malawi [MWI]	0.0029	0.0027	0.3311
Mauritius [MUS]	0.0036	0.0005	0.3300
Morroco [MAR]	0.0326	0.0055	0.3687
Nigeria [NGA]	0.0749	0.0291	0.2679
Sierra Leone [SLE]	0.0034	0.0012	0.0605
Zambia [ZMB]	0.0052	0.0019	0.7697
Zimbabwe [ZWE]	0.0080	0.0028	2.0604
Canada [CAN]	0.3097	0.0107	1.0003
Dominican Republic [DOM]	0.0109	0.0016	0.5427
Guatemala [GTM]	0.0145	0.0019	0.4386
Honduras [HND]	0.0048	0.0010	0.9482
Jamaica [JAM]	0.0052	0.0009	0.7455
Mexico [MEX]	0.3189	0.0198	0.7329
Panama [PAN]	0.0056	0.0006	1.5129
United States [USA]	3.3263	0.1028	0.8214
Argentina [ARG]	0.1529	0.0101	0.6399
Bolivia [BOL]	0.0094	0.0017	1.0253
Chile [CHL]	0.0391	0.0036	0.6451
Colombia [COL]	0.0684	0.0079	1.2511
Ecuador [ECU]	0.0203	0.0024	1.3615
Paraguay [PRY]	0.0057	0.0010	0.1050
Peru [PER]	0.0450	0.0050	0.9691
Hong Kong [HKG]	0.0395	0.0028	0.8568
India [IND]	0.6089	0.2585	0.5957
Israel [ISR]	0.0272	0.0014	0.9778
Japan [JPN]	1.0749	0.0706	1.2104
South Korea [KOR]	0.1177	0.0139	1.0065
Philippines [PHL]	0.0751	0.0169	0.8016
Sri Lanka [LKA]	0.0226	0.0051	1.6131
Syria [SYR]	0.0298	0.0021	0.9239
Taiwan [OAN]	0.0724	0.0069	1.1405
Thailand [THA]	0.0934	0.0223	0.6947
Austria [AUT]	0.0714	0.0034	0.9997
Belgium [BEL]	0.0980	0.0039	1.0820
Denmark [DNK]	0.0562	0.0026	1.1488
Finland [FIN]	0.0481	0.0023	1.5032
France [FRA]	0.5797	0.0233	1.0026
Germany [DEU]	0.6724	0.0273	1.1634
Greece [GRC]	0.0499	0.0036	1.1848
Iceland [ISL]	0.0023	0.0001	0.6380
Ireland [IRL]	0.0206	0.0012	0.9040
Italy [ITA]	0.5131	0.0218	0.9517
Luxembourg [LUX]	0.0042	0.0001	1.3084
Netherlands [NLD]	0.1450	0.0053	0.9271
Norway [NOR]	0.0445	0.0018	1.8028
Portugal [POR]	0.0434	0.0040	0.7004
Spain [ESP]	0.2568	0.0128	0.7997
Sweden [SWE]	0.0990	0.0040	1.0824
Switzerland [CHE]	0.0875	0.0031	1.8307
Turkey [TUR]	0.1222	0.0189	0.8257
United Kingdom [GBR]	0.5645	0.0267	0.7076
Yugoslavia [YUG]	0.0966	0.0098	0.5434
Australia [AUS]	0.1731	0.0064	1.1050
New Zealand [NZL]	0.0321	0.0013	1.0474

Sources : *Penn World Tables 5.6*. Ce tableau donne le produit réel ($\div 10^9$) moyen sur la période 1965–1990, le nombre de travailleurs ($\div 10^8$) moyen sur la même période et le ratio moyen du produit réel au stock de capital physique.

Tab. 3.2 – Description des données (Croissance)

	g_y	g_k	g_L
Ivory Coast [CIV]	0.0056	0.0533	0.0272
Kenya [KEN]	0.0122	−0.0071	0.0414
Madagascar [MDG]	−0.0140	0.0073	0.0210
Malawi [MWI]	0.0147	0.0586	0.0258
Mauritius [MUS]	0.0182	0.0265	0.0220
Morroco [MAR]	0.0171	0.0119	0.0305
Nigeria [NGA]	0.0137	0.0404	0.0253
Sierra Leone [SLE]	−0.0024	0.0493	0.0145
Zambia [ZMB]	−0.0164	−0.0217	0.0300
Zimbabwe [ZWE]	0.0043	−0.0150	0.0376
Canada [CAN]	0.0176	0.0342	0.0221
Dominican Republic [DOM]	0.0168	0.0536	0.0295
Guatemala [GTM]	0.0101	0.0219	0.0258
Honduras [HND]	0.0083	0.0170	0.0325
Jamaica [JAM]	−0.0014	−0.0008	0.0214
Mexico [MEX]	0.0157	0.0304	0.0311
Panama [PAN]	0.0114	0.0356	0.0274
United States [USA]	0.0109	0.0277	0.0170
Argentina [ARG]	0.0018	0.0286	0.0106
Bolivia [BOL]	0.0114	0.0309	0.0222
Chile [CHL]	0.0062	0.0257	0.0224
Colombia [COL]	0.0212	0.0249	0.0256
Ecuador [ECU]	0.0240	0.0405	0.0261
Paraguay [PRY]	0.0198	0.0617	0.0302
Peru [PER]	−0.0070	0.0122	0.0269
Hong Kong [HKG]	0.0515	0.0259	0.0266
India [IND]	0.0239	0.0369	0.0196
Israel [ISR]	0.0252	0.0229	0.0272
Japan [JPN]	0.0461	0.0804	0.0103
South Korea [KOR]	0.0685	0.0899	0.0241
Philippines [PHL]	0.0146	0.0147	0.0254
Sri Lanka [LKA]	0.0219	0.0310	0.0184
Syria [SYR]	0.0297	0.0176	0.0305
Taiwan [OAN]	0.0590	0.0903	0.0253
Thailand [THA]	0.0442	0.0664	0.0273
Austria [AUT]	0.0271	0.0598	0.0047
Belgium [BEL]	0.0234	0.0355	0.0059
Denmark [DNK]	0.0133	0.0346	0.0100
Finland [FIN]	0.0273	0.0405	0.0074
France [FRA]	0.0234	0.0435	0.0087
Germany [DEU]	0.0224	0.0395	0.0046
Greece [GRC]	0.0338	0.0523	0.0052
Iceland [ISL]	0.0206	0.0423	0.0214
Ireland [IRL]	0.0344	0.0467	0.0077
Italy [ITA]	0.0316	0.0394	0.0045
Luxembourg [LUX]	0.0234	0.0260	0.0080
Netherlands [NLD]	0.0167	0.0319	0.0140
Norway [NOR]	0.0214	0.0099	0.0147
Portugal [POR]	0.0403	0.0513	0.0089
Spain [ESP]	0.0305	0.0631	0.0074
Sweden [SWE]	0.0124	0.0372	0.0098
Switzerland [CHE]	0.0132	0.0355	0.0081
Turkey [TUR]	0.0337	0.0477	0.0195
United Kingdom [GBR]	0.0192	0.0360	0.0049
Yugoslavia [YUG]	0.0256	0.0513	0.0083
Australia [AUS]	0.0143	0.0253	0.0219
New Zealand [NZL]	0.0029	0.0241	0.0167

Sources : *Penn World Tables 5.6*. Ce tableau donne les taux de croissance annuels moyens du produit réel par travailleur, du stock de capital capital physique par travailleur et de la force de travail.

Tab. 3.3 – Croissance de la PGF et ses décompositions (Estimateur statique)

	ΔTFP	ΔEFF$^{\mathcal{D}}$	ΔTECH$^{\mathcal{D}}$
Ivory Coast [CIV]	−0.0018	−0.0129	0.0051
Kenya [KEN]	0.0137	0.0189	0.0071
Madagascar [MDG]	−0.0173	−0.0245	0.0095
Malawi [MWI]	−0.0098	−0.0011	−0.0012
Mauritius [MUS]	0.0069	0.0012	*.NaN*
Morroco [MAR]	0.0133	0.0027	0.0121
Nigeria [NGA]	0.0134	0.0000	−0.0028
Sierra Leone [SLE]	−0.0335	0.0000	*.NaN*
Zambia [ZMB]	−0.0104	−0.0092	0.0111
Zimbabwe [ZWE]	0.0112	0.0158	−0.0019
Canada [CAN]	0.0122	0.0064	0.0063
Dominican Republic [DOM]	−0.0080	−0.0148	0.0073
Guatemala [GTM]	−0.0002	−0.0045	0.0059
Honduras [HND]	−0.0023	−0.0038	0.0015
Jamaica [JAM]	−0.0040	−0.0035	−0.0006
Mexico [MEX]	−0.0019	0.0043	−0.0104
Panama [PAN]	−0.0097	−0.0110	0.0004
United States [USA]	0.0011	0.0000	0.0063
Argentina [ARG]	−0.0166	−0.0084	−0.0081
Bolivia [BOL]	−0.0068	−0.0081	0.0022
Chile [CHL]	−0.0106	−0.0091	−0.0008
Colombia [COL]	0.0051	0.0115	−0.0045
Ecuador [ECU]	0.0003	−0.0014	0.0010
Paraguay [PRY]	−0.0027	0.0000	*.NaN*
Peru [PER]	−0.0150	−0.0101	−0.0035
Hong Kong [HKG]	0.0340	0.0321	0.0032
India [IND]	0.0137	0.0000	−0.0021
Israel [ISR]	0.0093	0.0137	−0.0050
Japan [JPN]	0.0007	−0.0088	−0.0107
South Korea [KOR]	0.0155	0.0082	−0.0079
Philippines [PHL]	0.0070	0.0145	−0.0064
Sri Lanka [LKA]	0.0019	0.0046	−0.0042
Syria [SYR]	0.0182	0.0178	0.0000
Taiwan [OAN]	0.0050	0.0014	−0.0058
Thailand [THA]	0.0178	0.0071	−0.0010
Austria [AUT]	−0.0089	−0.0060	−0.0049
Belgium [BEL]	0.0080	0.0081	0.0002
Denmark [DNK]	−0.0087	−0.0030	−0.0050
Finland [FIN]	0.0174	0.0152	0.0039
France [FRA]	0.0008	0.0014	−0.0013
Germany [DEU]	0.0079	0.0069	0.0009
Greece [GRC]	0.0020	0.0036	−0.0034
Iceland [ISL]	−0.0042	0.0000	*.NaN*
Ireland [IRL]	0.0073	0.0073	−0.0026
Italy [ITA]	0.0051	0.0110	−0.0085
Luxembourg [LUX]	0.0233	0.0000	0.0176
Netherlands [NLD]	−0.0031	0.0022	−0.0053
Norway [NOR]	0.0214	0.0094	0.0117
Portugal [POR]	0.0066	0.0050	−0.0020
Spain [ESP]	−0.0065	−0.0066	−0.0055
Sweden [SWE]	−0.0030	−0.0018	−0.0010
Switzerland [CHE]	0.0132	0.0017	0.0113
Turkey [TUR]	0.0071	0.0087	−0.0061
United Kingdom [GBR]	−0.0018	−0.0000	−0.0092
Yugoslavia [YUG]	−0.0027	−0.0054	−0.0049
Australia [AUS]	0.0142	0.0032	0.0111
New Zealand [NZL]	−0.0057	−0.0078	0.0040

Les trois colonnes donnent, dans l'ordre, les taux de croissance annuels moyens de la PGF, de l'efficience (estimée avec la spécification VRS) et de la technologie (estimée avec la spécification VRS). Ces taux de croissance annuels moyens sont obtenus en considérant la moyenne géométrique des 25 facteurs de croissance annuels entre 1965 et 1990.

Tab. 3.4 – Croissance de la PGF et ses décompositions (Estimateur dynamique)

	ΔTFP	ΔEFF$^{\mathcal{D}}$	ΔTECH$^{\mathcal{D}}$
Ivory Coast [CIV]	−0.0031	−0.0195	0.0141
Kenya [KEN]	0.0127	0.0080	0.0118
Madagascar [MDG]	−0.0174	−0.0255	0.0100
Malawi [MWI]	−0.0042	−0.0116	0.0130
Mauritius [MUS]	0.0081	−0.0037	0.0115
Morroco [MAR]	0.0111	0.0026	0.0098
Nigeria [NGA]	0.0114	−0.0089	0.0224
Sierra Leone [SLE]	−0.0153	−0.0127	0.0088
Zambia [ZMB]	−0.0121	−0.0116	0.0102
Zimbabwe [ZWE]	0.0110	0.0134	0.0013
Canada [CAN]	0.0163	0.0064	0.0100
Dominican Republic [DOM]	−0.0095	−0.0168	0.0075
Guatemala [GTM]	−0.0013	−0.0078	0.0071
Honduras [HND]	−0.0032	−0.0071	0.0044
Jamaica [JAM]	−0.0045	−0.0083	0.0038
Mexico [MEX]	−0.0045	−0.0061	0.0000
Panama [PAN]	−0.0125	−0.0131	0.0006
United States [USA]	0.0079	−0.0000	0.0185
Argentina [ARG]	−0.0160	−0.0161	0.0002
Bolivia [BOL]	−0.0082	−0.0102	0.0027
Chile [CHL]	−0.0092	−0.0109	0.0011
Colombia [COL]	0.0045	0.0041	0.0004
Ecuador [ECU]	−0.0030	−0.0037	0.0007
Paraguay [PRY]	0.0069	−0.0032	0.0184
Peru [PER]	−0.0146	−0.0151	0.0007
Hong Kong [HKG]	0.0339	0.0320	0.0018
India [IND]	0.0137	−0.0010	0.0000
Israel [ISR]	0.0103	0.0104	0.0003
Japan [JPN]	0.0037	−0.0089	0.0023
South Korea [KOR]	0.0126	0.0006	0.0014
Philippines [PHL]	0.0066	0.0033	0.0033
Sri Lanka [LKA]	0.0026	0.0004	0.0008
Syria [SYR]	0.0169	0.0160	0.0009
Taiwan [OAN]	0.0038	−0.0010	0.0006
Thailand [THA]	0.0170	−0.0043	0.0073
Austria [AUT]	−0.0030	−0.0065	0.0034
Belgium [BEL]	0.0145	0.0080	0.0067
Denmark [DNK]	−0.0004	−0.0040	0.0038
Finland [FIN]	0.0232	0.0151	0.0088
France [FRA]	0.0067	0.0013	0.0054
Germany [DEU]	0.0137	0.0068	0.0069
Greece [GRC]	0.0007	0.0007	0.0001
Iceland [ISL]	−0.0064	−0.0040	0.0068
Ireland [IRL]	0.0039	0.0041	0.0002
Italy [ITA]	0.0118	0.0098	0.0018
Luxembourg [LUX]	0.0233	−0.0000	0.0199
Netherlands [NLD]	0.0049	0.0010	0.0039
Norway [NOR]	0.0214	0.0094	0.0118
Portugal [POR]	0.0079	0.0036	0.0014
Spain [ESP]	−0.0073	−0.0089	0.0000
Sweden [SWE]	0.0041	−0.0019	0.0063
Switzerland [CHE]	0.0132	0.0016	0.0114
Turkey [TUR]	0.0060	−0.0035	0.0025
United Kingdom [GBR]	−0.0045	−0.0052	0.0000
Yugoslavia [YUG]	−0.0036	−0.0135	0.0021
Australia [AUS]	0.0142	0.0030	0.0112
New Zealand [NZL]	−0.0002	−0.0087	0.0083

Les trois colonnes donnent, dans l'ordre, les taux de croissance annuels moyens de la PGF, de l'efficience (estimée avec la spécification VRS) et de la technologie (estimée avec la spécification VRS). Ces taux de croissance annuels moyens sont obtenus en considérant la moyenne géométrique des 25 facteurs de croissance annuels entre 1965 et 1990.

Tab. 3.5 – Des statistiques sur la croissance (Estimateur statique)

	ΔTFP	ΔEFF$^{\mathcal{D}}$	ΔTECH$^{\mathcal{D}}$	Δy
moyenne	0.0033	0.0015	0.0001	0.0200
médiane	0.0019	0.0014	−0.0010	0.0176
min	−0.0173	−0.0245	−0.0107	−0.0164
max	0.0340	0.0321	0.0176	0.0685
écart-type	0.0110	0.0099	0.0065	0.0161

Tab. 3.6 – Des statistiques sur la croissance (Estimateur dynamique)

	ΔTFP	ΔEFF$^{\mathcal{D}}$	ΔTECH$^{\mathcal{D}}$	Δy
moyenne	0.0038	−0.0021	0.0058	0.0196
médiane	0.0041	−0.0032	0.0038	0.0182
min	−0.0174	−0.0255	0.0000	−0.0164
max	0.0339	0.0320	0.0224	0.0685
écart-type	0.0111	0.0099	0.0057	0.0158

Tab. 3.7 – Nations sur la frontière

Années	Nations
1965	SLE, USA, ARG, PRY
1966	USA
1967	USA, PRY
1968	USA, PRY
1969	USA, PRY
1970	SLE, PRY
1971	USA, PRY
1972	USA, PRY
1973	USA, PRY
1974	PRY
1975	PRY
1976	PRY
1977	PRY
1978	USA, PRY
1979	USA
1980	PRY
1981	PRY
1982	$\emptyset$
1983	$\emptyset$
1984	USA
1985	USA
1986	USA
1987	USA
1988	HKG, USA
1989	USA, LUX
1990	HKG, LUX

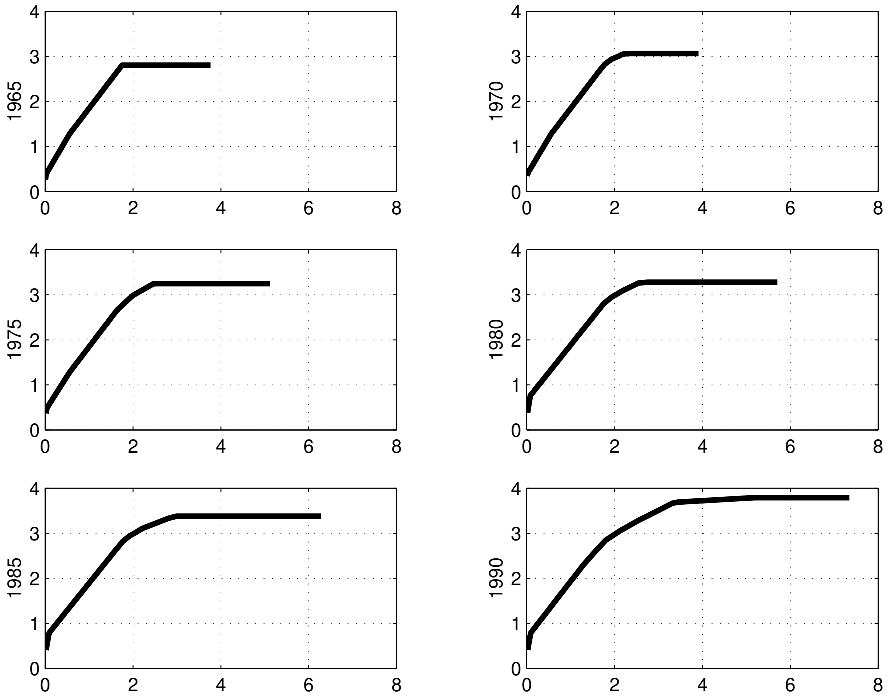

Fig. 3.11 – **Estimations de la frontière technologique mondiale**. Ces estimations sont obtenues avec la spécification CRS de la technologie.

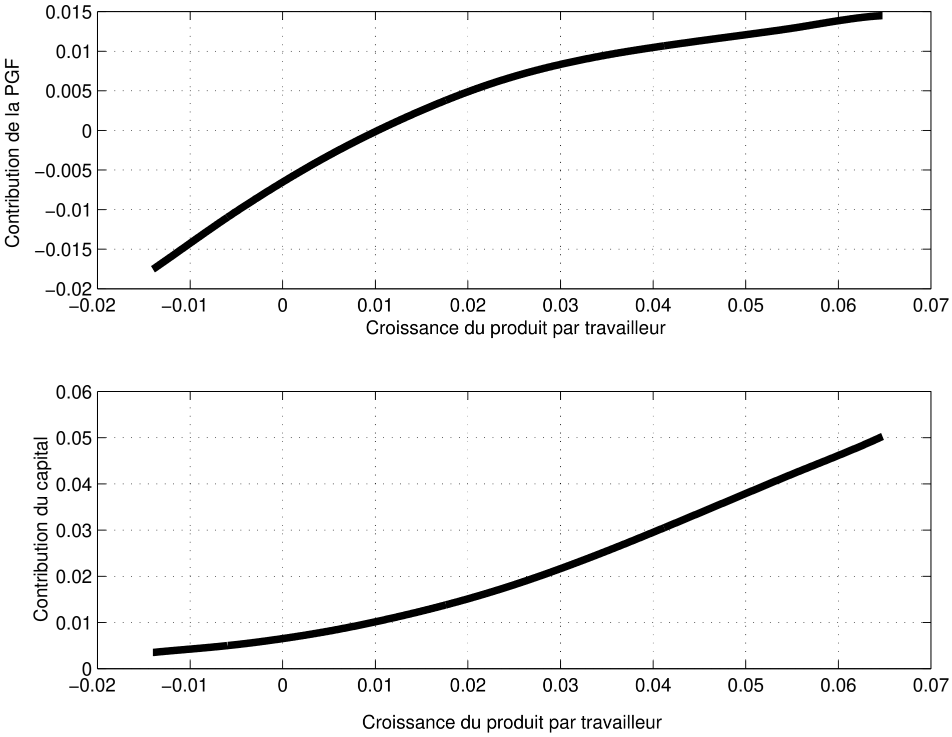

FIG. 3.12 – **Contribution locale à la croissance du produit par travailleur**. Dans le premier cadre la courbe noire représente la contribution locale de la croissance de la PGF à la croissance du produit par travailleur. Par complémentarité, la courbe noire du deuxième cadre révèle la contribution locale du capital par travailleur. Ces courbes sont obtenues en utilisant une régression non linéaire (*local polynomial regression*, voir Fan et Gijbels [54]).

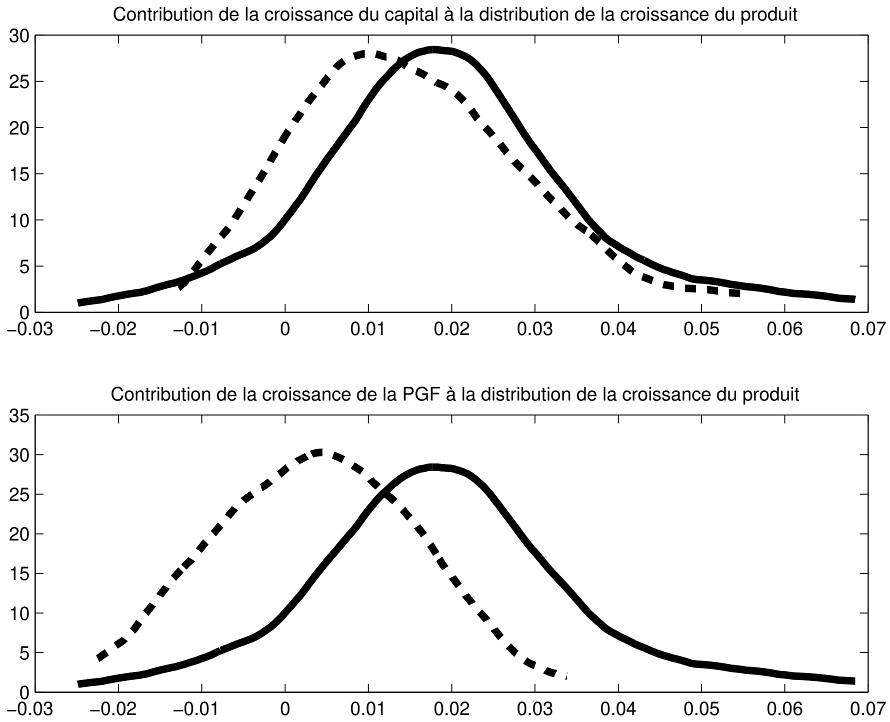

Fig. 3.13 – **Contributions à la distribution de la croissance du produit par travailleur**. La courbe pleine est un estimateur à noyau de la densité du taux de croissance du produit par travailleur (le paramètre de lissage est choisi par une procédure de *cross-validation*). La courbe en tirets est un estimateur à noyau de la densité du taux de croissance du produit par travailleur induit par l'accumulation du capital ou la dynamique de la productivité globale des facteurs. On a choisi le même paramètre de lissage pour tous les estimateurs.

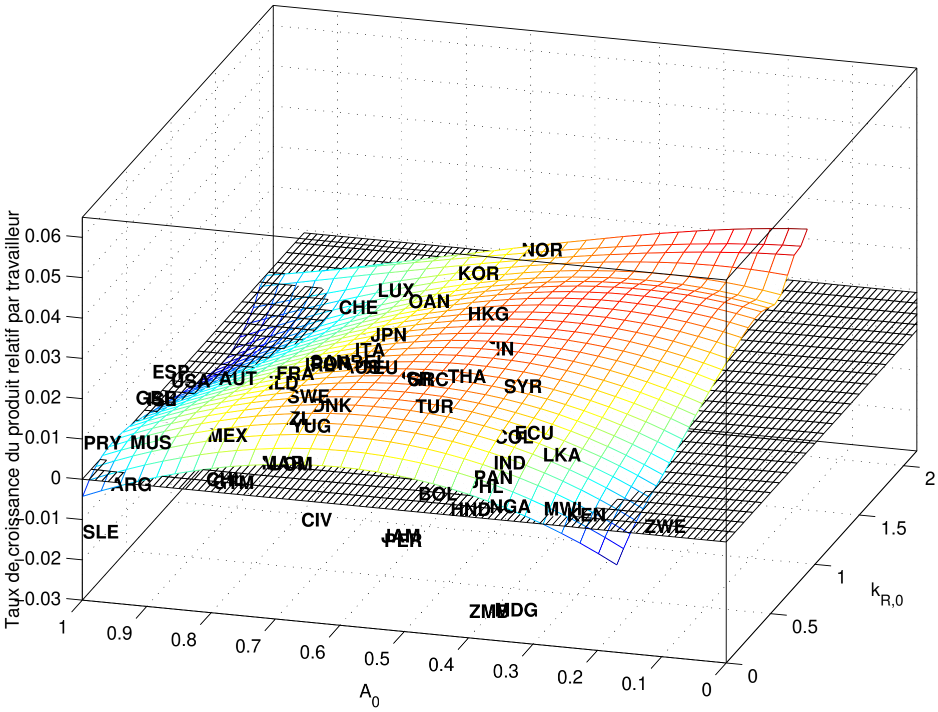

FIG. 3.14 – **Sources de la convergence (II)**. Cette surface est l'estimateur de Nadaraya-Watson de $m(\bullet, \bullet)$ dans la régression $g_{y,i} = m(k_{i,0}, A_{i,0}) + u_i$.

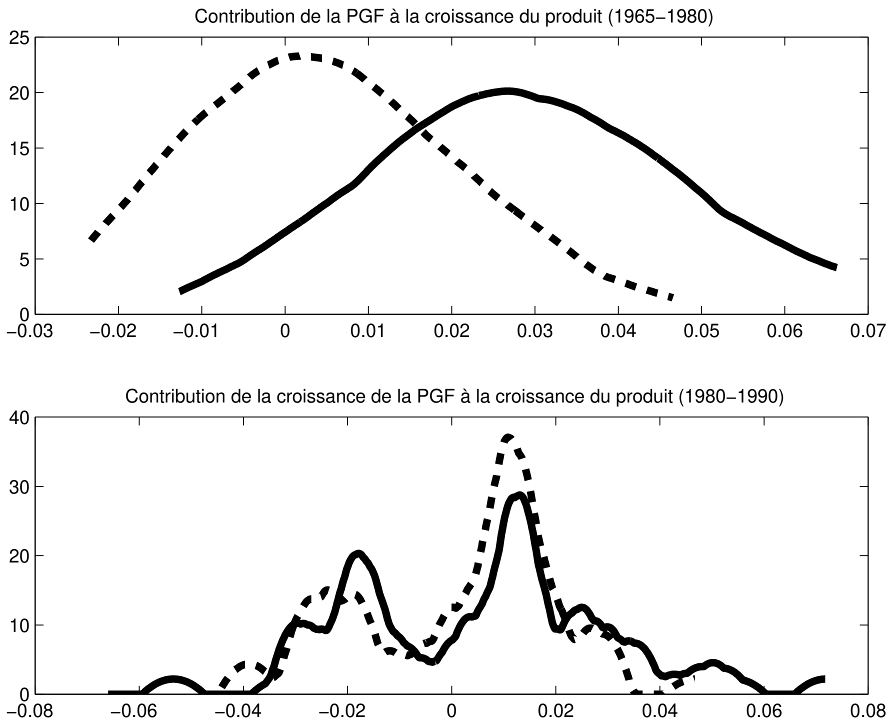

Fig. 3.15 – **Hétérogénéité chronologique de la contribution de la PGF à la croissance du produit par travailleur.** Les courbes noires représentent des estimateurs à noyau de la distribution du taux de croissance annuel moyen du produit par travailleur. Les courbes rouges représentent des estimateurs à noyau de la distribution du taux de croissance du produit par travailleur directement induit par la croissance de la PGF. Pour chaque période on a choisi un paramètre de lissage optimal avec une procédure de *cross-validation*.

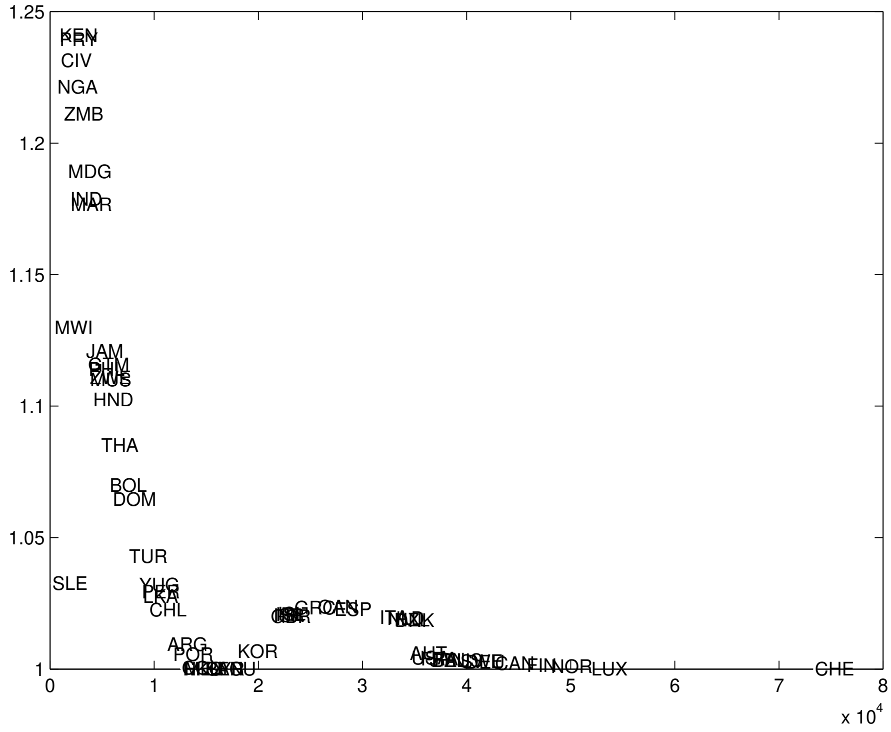

Fig. 3.16 – **Estimation dynamiquement cohérente de la frontière et biais**. Ce graphique représente les ratios $\widehat{D}_{HR,1990}/\widehat{D}_{1990}$ des distances à la frontière technologique estimée en 1990. $\widehat{D}$ correspond à notre estimation obtenue en considérant toutes les observations entre 1965 et 1990, et $\widehat{D}_{HR}$ correspond à l'estimation de Henderson et Russell obtenue en ne considérant que les observations en 1965 et 1990. Si ce ratio est strictement supérieur à l'unité, alors la distance estimée dans ce chapitre est plus grande que celle estimée par Henderson et Russell.

Tab. 3.8 – Estimations des frontières mondiales DEA et FDH en 1995

Nations	y	K/Y	$\widehat{\mathcal{D}}_{DEA}$	$\widehat{\mathcal{D}}_{FDH}$
AGO	2847	2.06107	0.40179	0.86746
ARG	24990	2.99981	0.61347	0.73474
AUS	45844	3.35556	0.82221	0.82221
AUT	44318	3.37232	0.79484	0.79484
BDI	1561	0.99909	0.54720	0.78561
BEL	51392	2.88328	0.92171	0.92171
BEN	2406	1.04768	0.63365	1.00000
BFA	2039	1.14190	0.56512	1.00000
BGD	6449	1.07402	0.79365	1.00000
BOL	6875	1.91469	0.49848	0.52917
BRA	18881	2.85851	0.56653	0.73636
BRB	27471	2.01190	0.81311	1.00000
BWA	16677	2.00939	0.63879	0.65040
CAF	2459	1.06066	0.63327	1.00000
CAN	44909	3.35207	0.80544	0.80544
CHE	44382	4.43517	0.79599	0.79599
CHL	22480	1.95081	0.75545	0.87672
CHN	4658	1.68861	0.51478	0.72228
CIV	4864	1.41719	0.60100	1.00000
CMR	3716	1.86840	0.45639	0.57621
COG	3338	3.63777	0.25372	0.25693
COL	12566	2.02963	0.58508	0.84290
COM	4532	1.37001	0.61072	1.00000
CPV	10454	1.67324	0.63405	0.70123
CRI	14175	2.56858	0.52248	0.55283
CYP	35542	2.35830	0.81036	1.00000
DNK	44613	3.19637	0.80013	0.80013
DOM	12232	1.81512	0.62983	0.82050
ECU	12511	3.11395	0.44635	0.48793
EGY	12992	0.91434	1.00000	1.00000
ESP	38413	3.00501	0.72417	0.74558
ETH	1343	0.68434	0.60402	0.67589
FIN	39206	4.03280	0.70316	0.70316
FJI	15890	2.21142	0.59554	0.61971
FRA	45875	3.34093	0.82277	0.82277
GAB	16259	2.43545	0.57544	0.63410
GBR	39497	2.47414	0.80993	1.00000
GHA	2883	1.71289	0.46708	0.90603
GIN	5771	1.73045	0.51840	0.79920
GMB	2638	1.00309	0.67290	1.00000

A suivre...

TAB. 3.9 – Estimations des frontières mondiales DEA et FDH en 1995 (suite)

Nations	y	K/Y	$\widehat{\mathcal{D}}_{DEA}$	$\widehat{\mathcal{D}}_{FDH}$
GNB	1994	2.19537	0.35471	0.62665
GRC	31042	3.28697	0.61729	0.78593
GTM	13983	1.18617	0.87814	0.93795
GUY	7890	3.31191	0.36081	0.48647
HKG	51521	2.04828	1.00000	1.00000
HND	7089	2.00529	0.49069	0.54564
IDN	9301	1.58360	0.62986	0.71590
IND	5287	1.31174	0.64996	0.81982
IRL	44871	2.34707	0.87220	1.00000
IRN	17138	2.94367	0.53416	0.66838
ISL	38032	3.66677	0.68210	0.68210
ISR	43680	3.29774	0.78340	0.78340
ITA	50244	2.99550	0.90112	0.90112
JAM	8136	3.18853	0.37407	0.50163
JOR	19493	1.96954	0.70044	0.76023
JPN	35898	3.86512	0.64383	0.64383
KEN	2696	1.82835	0.43741	0.84727
KOR	33070	2.86275	0.69345	0.93045
LKA	8013	1.26756	0.70912	1.00000
LSO	4742	1.47748	0.57800	0.73531
MAR	11598	1.96838	0.58542	0.77797
MDG	1968	0.58139	0.80504	0.99044
MEX	21690	3.39968	0.53818	0.63772
MLI	1873	1.58976	0.44112	0.71001
MOZ	1987	0.43705	0.91413	1.00000
MRT	2915	1.42781	0.53884	0.91609
MUS	25641	1.25541	1.00000	1.00000
MWI	1645	1.90674	0.37368	0.62358
MYS	25276	2.13183	0.75912	0.98576
NAM	15634	2.91038	0.51532	0.60973
NER	1818	1.52042	0.45042	0.68916
NGA	2031	1.46083	0.47952	0.76990
NIC	7132	2.47938	0.42947	0.47840
NLD	45175	3.51336	0.81021	0.81021
NOR	50312	4.19639	0.90234	0.90234
NPL	3282	1.56382	0.51578	1.00000
NZL	34506	3.58973	0.63497	0.66975
PAK	7221	1.35115	0.66207	1.00000
PAN	15434	2.70830	0.53162	0.60193
PER	10619	4.34632	0.34738	0.41414

A suivre...

TAB. 3.10 – Estimations des frontières mondiales DEA et FDH en 1995 (suite et fin)

Nations	y	K/Y	$\widehat{\mathcal{D}}_{DEA}$	$\widehat{\mathcal{D}}_{FDH}$
PHL	8019	2.49886	0.44369	0.53790
PNG	7562	1.65338	0.56515	0.58205
PRT	30305	2.48266	0.74220	0.89101
PRY	12796	1.98051	0.59860	0.85833
ROM	10097	3.25787	0.39000	0.39378
RWA	1801	0.73108	0.68896	0.90639
SEN	3182	1.28150	0.59735	1.00000
SGP	46763	4.64819	0.83869	0.83869
SLV	14908	1.00765	0.99724	1.00000
SWE	40908	3.03619	0.75206	0.79401
SYC	22401	1.53591	0.84779	0.87364
SYR	16219	1.57724	0.75349	1.00000
TCD	2822	1.82476	0.44233	0.85984
TGO	2238	1.34934	0.52199	0.84837
THA	13346	2.84077	0.48249	0.52049
TTO	22211	1.85735	0.77016	0.86623
TUR	15087	2.26978	0.57222	0.58839
TWN	34012	1.66936	0.99109	1.00000
TZA	1106	4.48992	0.17840	0.34758
UGA	1826	0.28181	1.00000	1.00000
URY	19961	1.89390	0.72265	0.77848
USA	55757	2.38645	1.00000	1.00000
VEN	19423	3.65971	0.49340	0.57106
ZAF	21543	2.08449	0.71504	0.84018
ZAR	654	2.30273	0.23356	0.32914
ZMB	2600	4.01278	0.22469	0.32447
ZWE	5644	3.87705	0.29363	0.37859

Sources : *Penn World Tables* 6. Voir la section [5.1, en page 222] pour la construction des données relatives au stock de capital. La première colonne désigne les nations de l'échantillon par les codes utilisés dans les *Penn World Tables*. La deuxième colonne donne le produit réel par travailleur en 1995. La troisième colonne présente nos estimations du ratio capital–produit en 1995. Les dernières colonnes reportent nos estimations des distances aux frontières DEA et FDH. Pour la frontière DEA on a considéré la spécification CRS de la technologie.

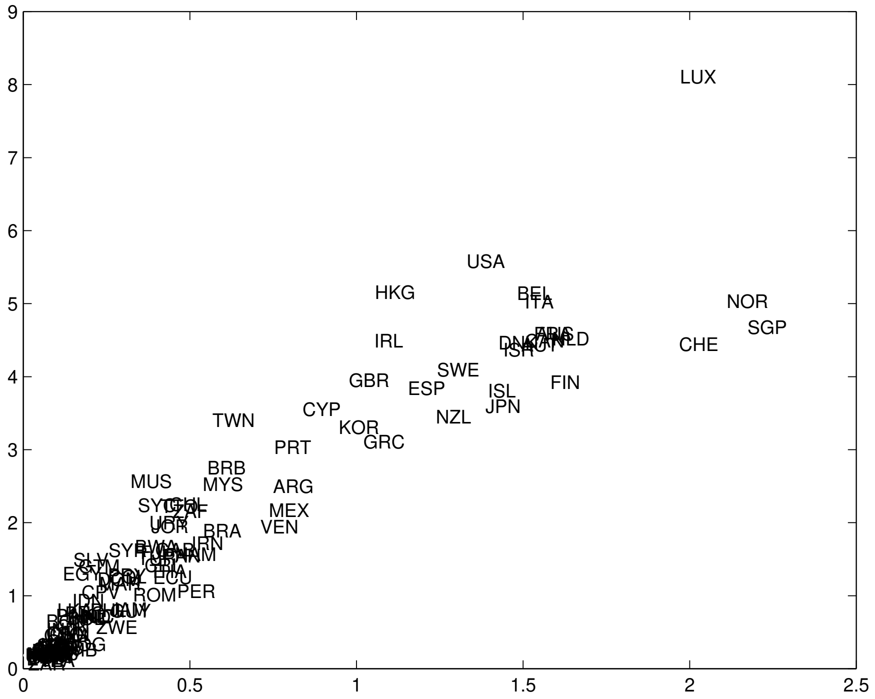

FIG. 3.17 – **Représentations des données dans le plan capital par travailleur – produit par travailleur.** Le capital par travailleur est exprimé en centaine de milliers de dollars, le produit par travailleur en dizaine de milliers de dollars. Voir la section [5.1, en page 222].

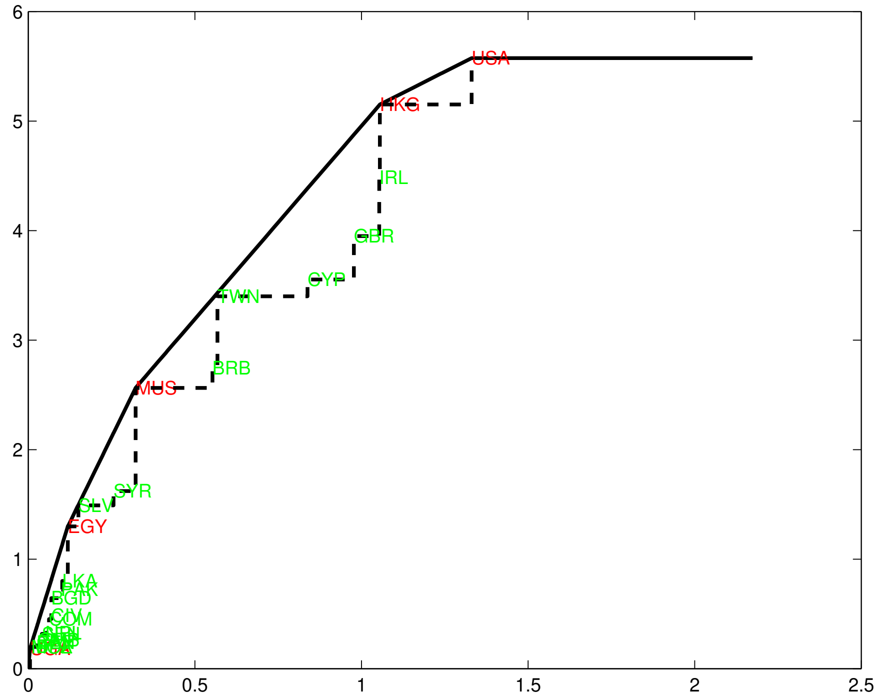

FIG. 3.18 – **Comparaison des frontières NIRS et FDH.** La courbe pleine correspond à la frontière NIRS. La courbe en tirets correspond à la frontière FDH.

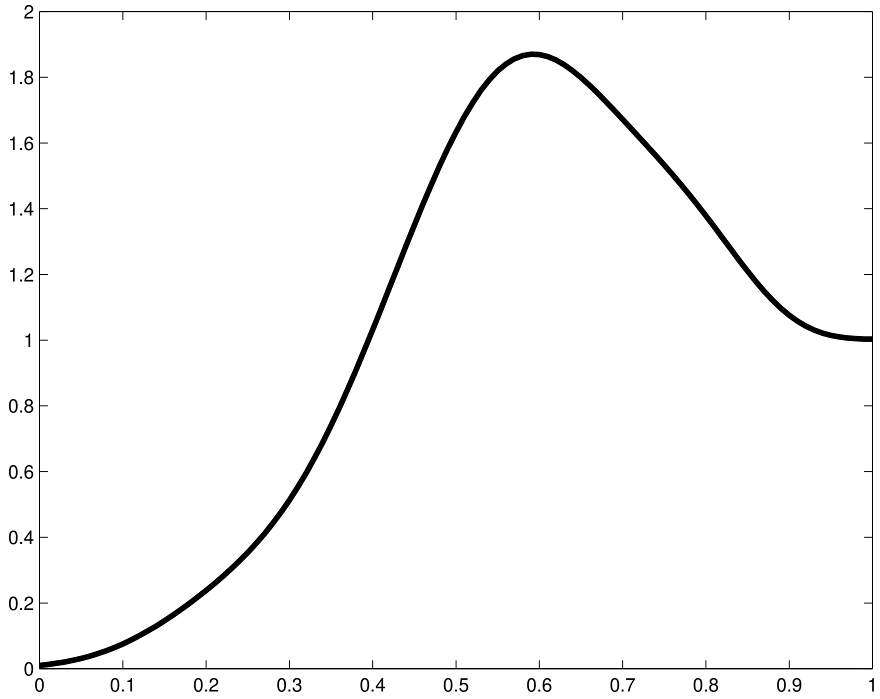

FIG. 3.19 – **Densité estimée des distances à la frontière NIRS.** On a utilisé un estimateur à noyau avec une réflexion sur la borne supérieure (voir Silverman [132, en pages 29-32]), le paramètre de lissage optimal ($h^* = 0.09$) est déterminé par une procédure de *cross-validation* (sans considérer les économies efficientes, voir Silverman [132, en pages 51-52]).

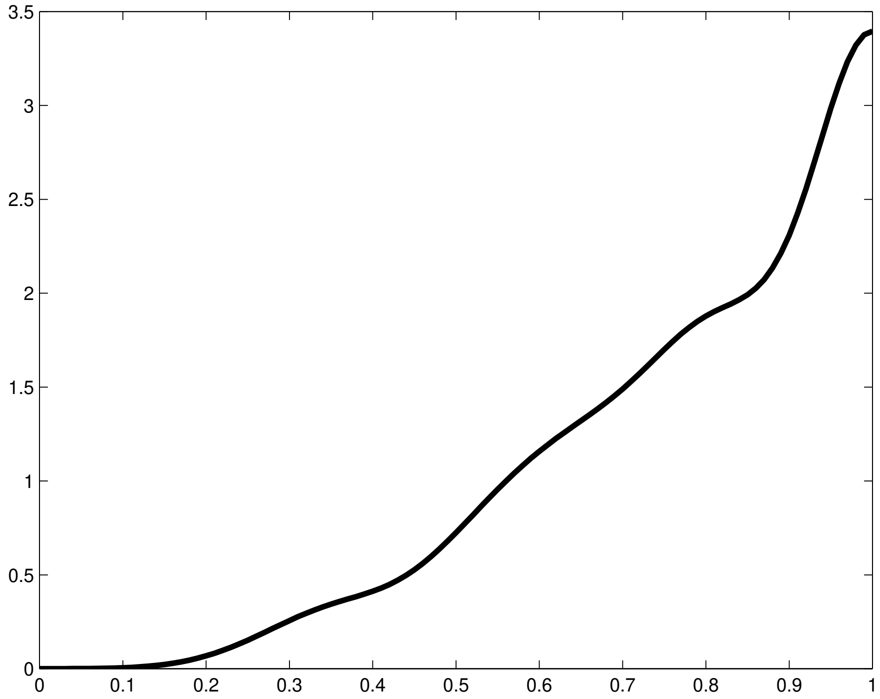

FIG. 3.20 – **Densité estimée des distances à la frontière FDH.** On a utilisé un estimateur à noyau avec une réflexion sur la borne supérieure (voir Silverman [132, en pages 29-32]), le paramètre de lissage optimal ($h^* = 0.07$) est déterminé par une procédure de *cross-validation* (sans considérer les économies efficientes, voir Silverman [132, en pages 51-52]).

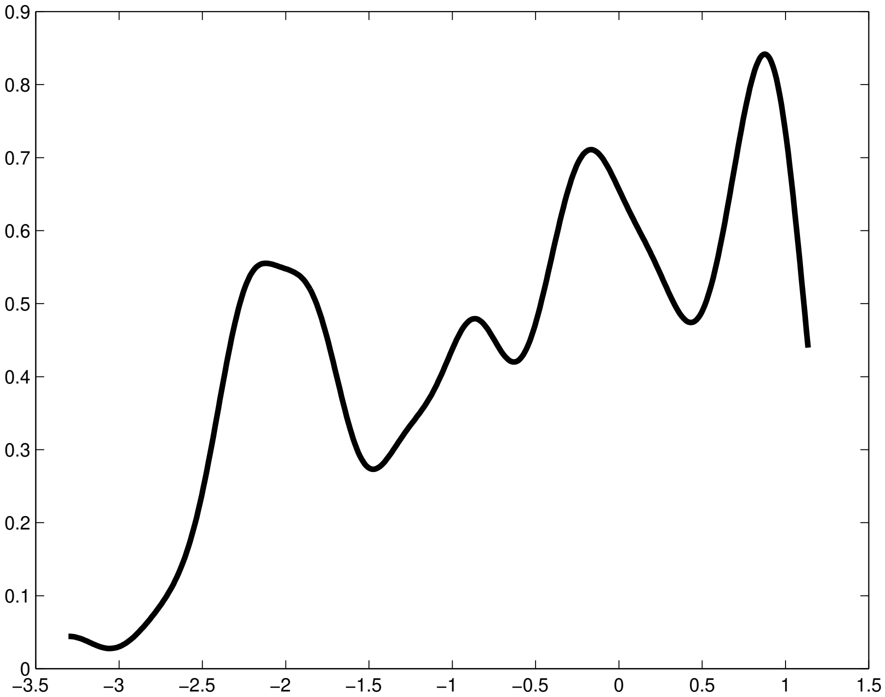

Fig. 3.21 – **Densité estimée du produit par travailleur relatif en 1995.** On a utilisé un estimateur à noyau, le paramètre de lissage optimal ($h^* = 0.17$) est déterminé par une procédure de *cross-validation*.

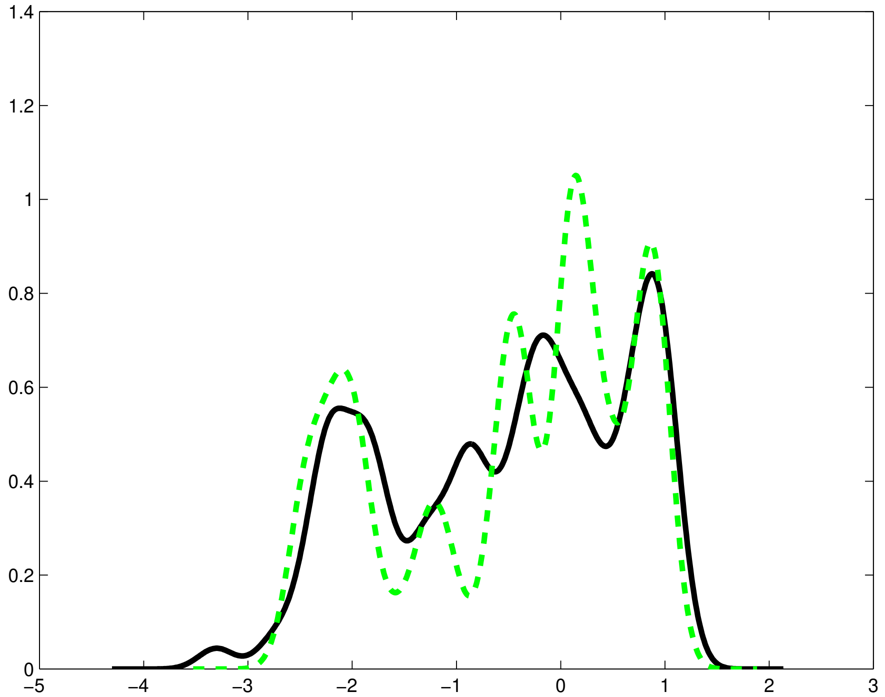

Fig. 3.22 – **Rôle de l'hétérogénéité relative à la dotation en capital sur les niveaux observés du produit par travailleur.** La courbe en trait plein correspond à un estimateur de la distribution du produit par travailleur observé. La courbe en tirets correspond à un estimateur de la distribution produit par travailleur lorsque les nations sont efficaces. On a utilisé des estimateurs à noyau avec un paramètre de lissage commun ($h = 0.17$).

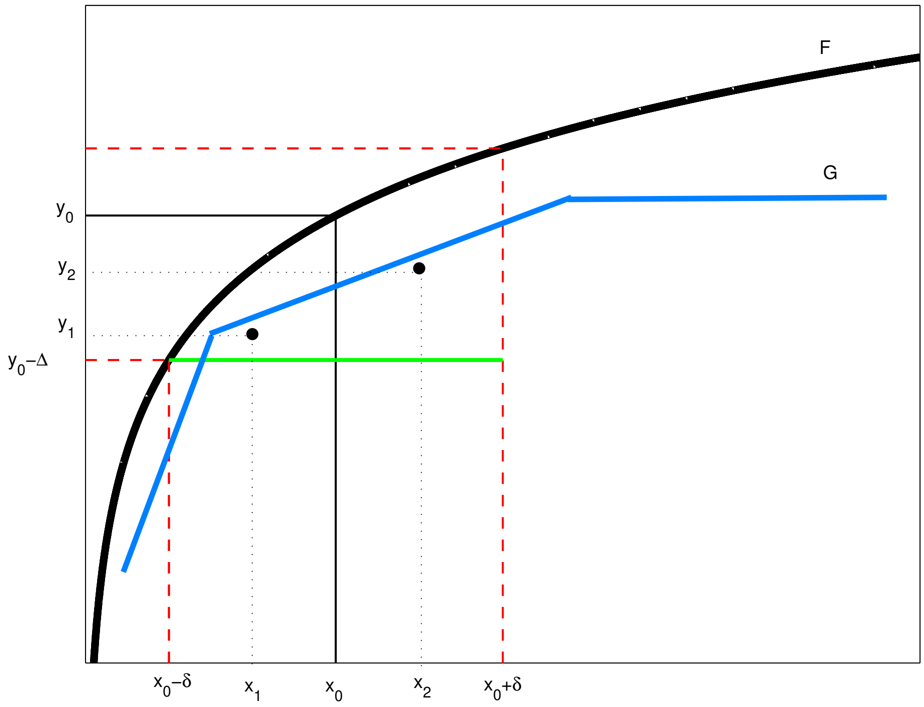

FIG. 3.23 – **Preuve de la proposition [2.1].** La courbe F correspond à la vraie frontière. La courbe G correspond à l'estimateur VRS de la frontière.

Conclusion générale

Dans cette thèse on s'est intéressé aux rôles respectifs de l'accumulation de capital physique, du progrès technologique et de l'efficience dans la dynamique mondiale de croissance et dans la convergence ou divergence des régions ou nations.

Les résultats obtenus contribuent à relativiser le rôle de la technologie. Sur des données régionales américaines on a vu, en détournant des régressions à la Barro, que la β-convergence est essentiellement expliquée par l'histoire néoclassique. Pour les états américains, la condition initiale du produit par tête rend plus compte de la variabilité initiale du stock de capital physique que de la variabilité initiale des niveaux technologiques. Ainsi la β-convergence est plus le résultat d'un *capital deepening effect* que d'un *technological catch-up*. Sur un panel international, un exercice de comptabilité de la croissance nous permet d'obtenir une conclusion analogue. Les contributions à la croissance du produit par travailleur du progrès technologique et de la dynamique d'efficience sont généralement faibles. Les nations composants cet échantillon divergent (ce résultat est établi dans le chapitre [2]). Il apparaît que le moteur de cette divergence réside essentiellement dans la dynamique non linéaire du capital physique, qui pourrait s'expliquer par la non convexité de l'ensemble de production mondial.

Il n'est pas innocent que des économistes comme Easterly et Levine [48], qui travaillent à la Banque Mondiale, affirment que la croissance mondiale est essentiellement déterminée par la Productivité Globale des Facteurs. En effet, on ne conseillera pas la même politique de développement si la diversité des performances en termes de croissance s'explique par l'accumulation de facteurs ou l'imparfaite diffusion des technologies. Ces deux auteurs biaisent leur conclusion en faveur de la technologie en se limitant à la période 1980-1992. Si on remonte plus loin dans le passé, il apparaît clairement que les miracles économiques ne sont pas le fruit d'une plus grande ouverture facilitant la diffusion des technologies mais le résultat d'une dynamique soutenue d'accumulation du capital physique. Mettre en évidence le rôle important de la technologie dans les années 80, comme Easterly et Levine, n'est pas très intéressant. En effet, nous avons vu dès l'introduction générale que le mouvement de divergence des nations se manifeste jusque dans les années 70. A partir des années 80 la distribution du produit par tête n'évolue plus, les miracles et désastres économiques ont déjà formé la distribution du produit par tête.

Le contenu de la thèse est limité dans plusieurs directions qui appellent de nombreux prolongements. Commençons par le premier chapitre. L'interprétation d'une régression à la Barro est évidemment conditionnelle au modèle considéré. Le problème du travail présenté dans le premier chapitre est que, même si notre interprétation fournit un éclairage original sur le débat entre convergence néo-classique et rattrapage technologique, le modèle à la base de cette interprétation peut sembler beaucoup trop simple. On pourra toujours montrer que la vitesse de convergence mesurée par une régression à la Barro est instable si la dynamique de productivité est endogène, car pour obtenir cette instabilité il suffit que le processus générateur des données contienne plus d'une variable d'état. Cependant, il n'est pas certain que les trajectoires de la vitesse de convergence soient aussi simples. Ceci reste à vérifier, s'il est encore possible d'étudier le problème de façon analytique. L'exogénéité de la dynamique de rattrapage technologique n'est pas le seul problème. La fiction d'un monde (ou d'une nation) composé de pays (régions) autarciques trouve peu de justifications. La fiction alternative d'un monde intégré sans frottements n'est pas plus défendable. Un modèle intermédiaire satisfaisant est difficile à imaginer et on ne peut espérer obtenir des résultats analytiques de ce côté.
La démarche proposée pour estimer les vitesses de convergence a le mérite d'être cohérente avec un modèle et d'être originale par rapport au reste de la littérature. Pour autant cet exercice se révèle insatisfaisant pour deux raisons. Premièrement on en demande sûrement trop au modèle. Ceci rejoint

la remarque précédente. Le modèle envisagé est probablement trop "stylisé" pour que l'on puisse accorder beaucoup de crédit à ces estimations. On sait très bien par ailleurs, *ie* la littérature des cycles réels, que ce simple modèle est incapable de reproduire un grand nombre des faits stylisés de la dynamique de court terme. Deuxièmement, l'estimation du processus ARIMA sur le produit par travailleur américain est assez décevante. L'estimation d'une racine unitaire dans la partie MA nous empêche d'évaluer le bien fondé de la spécification retenue et donc du modèle. Ainsi, l'interprétation des racines de la partie autorégressive comme les vitesses de convergence néo-classique et technologique doit être relativisée.

Dans le deuxième chapitre nous sommes revenu sur l'approche en séries temporelles des tests de la convergence. A mon sens, il s'agit de l'approche la plus intéressante pour tester l'hypothèse de convergence, même si elle n'a pas rencontré le même succès que l'*approche classique*. Néanmoins elle souffre encore d'importantes limites.
Dans ce chapitre on a premièrement établi les limites de certaines procédures de test en panel. Les procédures reposant sur une écriture à la Dickey-Fuller du processus générateur des données, comme dans Evans et Karras [52], ne sont pas satisfaisantes pour tester l'hypothèse de convergence des nations. En employant une représentation de la partie déterministe plus cohérente, on a mis en évidence les limites de la conclusion de ces deux auteurs en faveur de la convergence pour une large coupe de nations. Ces auteurs concluent aussi, en employant la même stratégie de test de racine unitaire en panel, la convergence des états américains. Nous ne parviendrions vraisemblablement pas à rejeter la divergence en appliquant le test de Choi [30] en panel et le test de Schmidt et Phillips [130] sur cet échantillon. Comme le soulignent Bernard et Durlauf [21], il s'agit de la limite la plus importante de l'approche en séries temporelles. Si les données sont caractérisées par une forte dynamique de transition un test de racine unitaire conclut généralement à la divergence, même si cette dynamique de transition résulte en un rapprochement des économies. En effet, lorsque l'écart initial du produit par tête à sa cible de long terme est trop important, au sens où les données ne sont pas cohérentes avec une linéarisation de la dynamique d'un modèle de croissance autour de l'état stationnaire, le retour à l'état stationnaire n'est

pas géométrique et une procédure de test de racine unitaire tend à assimiler cette distance à un choc permanent en concluant à la divergence. Pour dépasser cette limite on a proposé d'associer la dynamique de transition à la partie déterministe du processus. On a utilisé une version multivariée du test de Kwiatkowski, Phillips, Schmidt et Shin [86] en autorisant la présence d'une rupture commune sur la tendance linéaire caractérisant le produit par travailleur relatif. Sur une première période, durant la phase de rattrapage, la pente de la tendance linéaire est spécifique à chaque nation. Durant une deuxième période, qui suit la phase de rattrapage, le produit par travailleur relatif est stationnaire en niveau. On propose une statistique pour tester la stationnarité stochastique, autour d'une tendance linéaire sur une première période puis autour d'un niveau (éventuellement nul) sur une seconde période, des processus de produit par travailleur relatif. Après avoir décrit le comportement asymptotique de cette statistique de test, on l'a appliqué sur un échantillon de dix nations européennes. On ne parvient pas à rejeter la convergence absolue contre la divergence ou la convergence conditionnelle.

Ce travail est limité dans au moins deux directions. Premièrement, la statistique proposée ne pourrait être considérée pour tester la stationnarité d'une large coupe de nations ou régions. Il n'y a pas d'asymptotique possible sur la dimension individuelle, qui ne peut être supérieure à la dimension chronologique (la matrice de variance-covariance de long terme ne serait plus inversible). Cette limite explique la taille de l'échantillon européen considéré. On ne peut donc appliquer cette stratégie sur l'échantillon des états américains. Pour dépasser cette limite il faudrait plutôt utiliser une stratégie en panel. Dans cette perspective on pourrait reprendre l'approche proposée par Bai et Ng [9] en autorisant des ruptures dans la partie déterministe. Cela nous permettrait d'avoir une asymptotique dans les dimensions individuelle et chronologique, tout en contrôlant de la dépendance spatiale.

Deuxièmement, la représentation de la dynamique de transition par une tendance linéaire segmentée est finalement assez discutable. On pourrait envisager des spécifications plus ou moins complexes de la partie déterministe pour rendre compte de la dynamique de transition, mais elles seraient toutes aussi difficilement justifiables que la tendance linéaire segmentée. Il serait plus satisfaisant d'estimer la part de la dynamique des écarts de produit par tête associée à la transition en utilisant un modèle de croissance, puis de

tester la stationnarité de la dynamique résiduelle. Une telle approche, si elle est envisageable, reste à construire.
Enfin, on peut penser que la mise en œuvre de tests, toujours plus perfectionnés, pour tester la convergence des nations sur de très larges coupes n'a plus grand intérêt aujourd'hui (à part un intérêt purement économétrique, non négligeable, en proposant des procédures de tests de racine unitaire ou de cointégration toujours plus robustes). La divergence mondiale des nations est aujourd'hui généralement admise. Un problème plus intéressant, en suivant les stratégies proposées par Bai et Ng [9] et [10] serait d'établir selon quel schéma ces nations divergent. L'identification des tendances communes de la dynamique mondiale du produit par tête pourrait nous permettre, au-delà d'un simple constat, de mieux comprendre les sources de la divergence des nations.

Dans le dernier chapitre nous avons mis en œuvre un exercice de *growth accounting* pour évaluer, de façon plus traditionnelle que dans le premier chapitre, les contributions respectives de la technologie et de l'accumulation du capital physique. L'approche envisagée dans ce chapitre est limitée dans de multiples directions.
Commençons par les remarques d'ordre technique. Les résultats présentés dans la troisième section du chapitre [3] sont donnés sans intervalles de confiance. On ne peut dire par exemple si les taux de croissance de la PGF reportés sont significativement différents de zéro. Une procédure de *bootstrap*, cohérente avec notre approche qui impose la croissance des ensembles de production, reste à construire. Une autre limite importante de cette approche est son ignorance des erreurs de mesure. Il est de plus difficile de concevoir dans quel sens cette ignorance peut affecter nos conclusions.
Nous n'avons laissé pratiquement aucune place au capital humain. La littérature empirique de la croissance n'est pas décisive sur le canal par lequel le capital humain doit jouer sur la croissance. Par exemple Benhabib et Spiegel [18] concluent que le capital humain n'influence pas la croissance en temps qu'*input* de la technologie mais via le A. Il serait intéressant de reprendre ce genre d'exercice de spécification dans le cadre d'une approche non paramétrique comme l'approche DEA.
Le choix de la direction pour évaluer la distance à la frontière technologique

mondiale pose aussi problème. Les conclusions pourraient être moins favorables au rôle prééminent du capital physique si nous considérions une mesure non radiale de la distance à la frontière, conforme avec une efficience Harrod neutre. Ceci reste à vérifier.

Enfin le test de convexité de l'ensemble de production mondial remet au cause l'approche employée pour estimer les indices de croissance de la PGF. Si cet ensemble est non convexe, on ne peut utiliser la frontière CRS (ou même VRS) pour dériver les indices de croissance de la PGF et ses décompositions. Il faudrait reprendre ces estimations en estimant des frontières FDH. Comme le notent Briec, Kerstens et Vanden Eeckaut [26], cet estimateur converge vers la vraie frontière quelle que soit la nature des rendements d'échelle. Dès lors, si on ne dispose pas d'un a priori fort en faveur d'une fonction de production concave, il est préférable d'utiliser l'estimateur FDH. Ce choix, bien sûr, implique une détérioration de la vitesse de convergence de l'estimateur de la frontière.

Une limite plus fondamentale de ce type d'exercice est son incapacité à tenir compte de l'endogénéité des variables étudiées. Affirmer la faiblesse de la contribution de la PGF en raisonnant *ceteris paribus* peut nous amener à sur-évaluer la contribution du capital. Cette littérature empirique manque de modèles.

Bibliographie

[1] Moses Abramovitz. Resource and output trends in the united states since 1870. *American Economic Review*, 46(2) :5–23, 1956.

[2] Moses Abramovitz. Catching-up, forging ahead and falling behind. *Journal of Economic History*, 46 :385–406, 1986.

[3] Stéphane Adjemian. Convergence des économies : une critique de la procédure d'Evans et Karras. mémoire de DEA, Université d'Evry - Val d'Essonne, Septembre 1998.

[4] Stéphane Adjemian. Convergence des productivités européennes : Réconcilier deux approches de la convergence. Document de recherche 99-17, EPEE, Université d'Evry - Val d'Essonne, Juillet 1999.

[5] Stéphane Adjemian. Convergence des productivités européennes : Transition, rupture et racine unitaire. mimeo, à paraître dans *Les Annales d'Économies et de Statistiques*, EPEE, Université d'Evry - Val d'Essonne, Janvier 2002.

[6] Stéphane Adjemian, Jérôme Glachant, et Charles Vellutini. Regional convergence and aggregate growth. mimeo, EPEE, Université d'Evry - Val d'Essonne, Janvier 2000.

[7] Byung Chul Ahn. Testing the null of stationarity in the presence of structural breaks for multiple time series. mimeo, Department of Economics, The Ohio State University, Colombus, août 1994.

[8] Donald W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59 :817–858, 1991.

[9] Jushan Bai et Serena Ng. Determining the number of factors in approximate factor models. mimeo, Department of Economics, Boston College, Décembre 2000.

[10] Jushan Bai et Serena Ng. A PANIC attack on unit roots and cointegration. mimeo, Department of Economics, Boston College, Décembre 2001.

[11] Rajiv D. Banker. Maximum likelihood, consistency and data envelopment analysis : A statistical foundation. *Management Science*, 39(10) :1265–1273, 1993.

[12] Robert J. Barro. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106 :407–43, 1991.

[13] Robert J. Barro et Jong-Wha Lee. International data on educational attainment updates and implications. Working Paper 7911, NBER, 2000.

[14] Robert J. Barro et Xavier Sala-i-Martin. Economic growth and convergence across the united states. Working paper 3419, NBER, 1990.

[15] Robert J. Barro et Xavier Sala-i-Martin. Convergence. *Journal of Political Economy*, 100(2) :223–251, 1992.

[16] Robert J. Barro et Xavier Sala-i-Martin. *Economic Growth*. McGraw Hill, New York, 1995.

[17] William J. Baumol. Productivity growth, convergence, and welfare : What the long run-data show. *The American Economic Review*, 76 :1072–1085, 1986.

[18] Jess Benhabib et Mark Spiegel. The role of human capital in economic development : Evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34 :143–173, 1994.

[19] Andrew B. Bernard et Stephen N. Durlauf. Convergence in international output. *Journal of Applied Econometrics*, 10 :97–108, 1995.

[20] Andrew B. Bernard et Stephen N. Durlauf. Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 85 :161–173, 1996.

[21] Andrew B. Bernard et Steven N. Durlauf. Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71 :161–173, 1996.

[22] Andrew B. Bernard et Charles Jones. Productivity across industries and countries : Time series theory and evidence. *Review of Economics and Statistics*, 78(1) :135–146, Février 1996.

[23] Andrew B. Bernard et Charles I. Jones. Comparing apples to oranges : productivity convergence and measurment across industries and countries. *The American Economic Review*, 86 :1216–1238, 1996.

[24] Marco Bianchi. Testing for convergence : Evidence from non-parametric multimodality tests. *Journal of Applied Econometrics*, 12(4) :393–409, Juillet 1997.

[25] Herman J. Bierens. Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the u.s. price level and interest rate. *Journal of Econometrics*, 81 :29–64, 1997.

[26] Walter Briec, Kristiaan Kerstens, et Phillipe Vanden Eeckaut. Nonconvex technologies and cost functions : Definitions, duality and nonparametric tests of convexity. Efficiency Series Paper 03/2002, Universidad de Oviedo, 2002.

[27] John Y. Campbell. Inspecting the mechanism : An analytical approach to the stochastic growth model. *Journal of Monetary Economics*, 33 :463–508, 1994.

[28] Gerald A. Carlino et Leonard O. Mills. Are u.s. regional incomes converging ? *Journal of Monetary Economics*, 32 :335–346, 1993.

[29] Robert G. Chambers. *Applied Production Analysis : the dual approach*. Cambridge University Press, 1988.

[30] In Choi. Unit root tests for panel data. *Journal of International Money and Finance*, 20 :249–272, 2001.

[31] In Choi et Byung Chul Ahn. Testing the null of stationarity for multiple time series. *journal of Econometrics*, 88 :41–77, 1999.

[32] T.G. Conley. GMM estimation with cross sectional dependence. *journal of Econometrics*, 92 :1–45, 1999.

[33] Angel De la Fuente. The emprics of growth and convergence : a selective review. *Journal of Economic Dynamics and Control*, 21(11) :23–73, 1997.

[34] D. Deprins, L. Simar, et D. Tulkens. *Measuring Labor Inefficiency in Post Offices*, pages 243–267. North-Holland, Amsterdam, 1984.

[35] Alain Desdoigts. *Changes in the World Income Distribution : A Nonparametric Approach to Challenge the Neoclassical Convergence Argument*. PhD thesis, European University Institute, Florence, Juin 1994.

[36] Alain Desdoigts. Patterns of economic development and the formations of clubs. *Journal of Economic Growth*, 4 :305–330, 1999.

[37] Alain Desdoigts. Neoclasical convergence versus technological catch-up : A contribution for reaching a consensus. Nota di Lavoro 8.2002, Fondazione Eni Enrico Mattei & EPEE, Université d'Évry, Janvier 2002.

[38] W. Erwin Diewert. Capital and the theory of productivity measurement. *American Economic Review*, 70 :260–267, 1980.

[39] Steve Dowrick et Duc Tho Nguyen. O.e.c.d. comparative economic growth 1950-85 : Catch-up and convergence. *American Economic Review*, 79(5) :1010–1030, 1989.

[40] Steve Dowrick et John Quiggin. International comparisons of living standards and tastes : A revealed-preference analysis. *American Economic Review*, 84(1) :332–341, 1994.

[41] Steve Dowrick et John Quiggin. True measures of gdp and convergence. *American Economic Review*, 87(1) :41–64, 1997.

[42] Steve Dowrick et Mark Rogers. Classical and technological convergence : beyond the solow-swan growth model. *Oxford Economic Papers*, 54(3) :369–385, 2002.

[43] John Duffy et Chris Papageorgiou. A cross country empirical investigation of the aggregate production function specification. *Journal of Economic Growth*, 5 :87–120, 2000.

[44] Stephen N. Durlauf. Econometric analysis and the study of economic growth : A skeptical perspective. mimeo, University of Wisconsin, Mai 2000.

[45] Stephen N. Durlauf et Paul A. Johnson. Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, 10 :365–384, 1995.

[46] Stephen N. Durlauf, Andros Kourtellos, et Arthur Minkin. The local solow growth model. mimeo, University of Wisconsin, Octobre 2000.

[47] Stephen N. Durlauf et Danny T. Quah. The new empirics of economic growth. Discussion Paper 384, CEP, Janvier 1998.

[48] William Easterly et Ross Levine. It's not factor accumulation : Stylized facts and growth models. Discussion paper, World Bank, Mars 2001.

[49] Paul Evans. Using cross-country variances to evaluate growth theories. *Journal of Economic Dynamics and Control*, 20 :1027–1049, 1996.

[50] Paul Evans. How do fast economies converge. *Review of Economics and Statistics*, 79 :219–225, 1997.

[51] Paul Evans. Using panel data to evaluate growth theories. *International Economic Review*, 39 :295–306, 1998 a.

[52] Paul Evans et Georgio Karras. Convergence revisited. *Journal of Monetary Economics*, 37 :249–265, 1996.

[53] Jan Fagerberg. Technology and international differences in growth rates. *Journal of Economic Litterature*, XXXII :1147–1175, 1994.

[54] J. Fan et I. Gijbels. *Local Polynomial Modelling and its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York, 1996.

[55] Rolf Färe, Shawna Grosskopf, et C.A. Knox Lovell. *Production Frontiers*. Cambridge University Press, 1994.

[56] Rolf Färe, Shawna Grosskopf, Mary Norris, et Zhongyang Zhang. Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1) :66–83, 1994.

[57] Rolf Färe et C.A. Knox Lovell. Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1) :150–162, 1978.

[58] M.J. Farell. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, 120(III) :253–281, 1957.

[59] Barbara M. Fraumeni. The measurement of depreciation in the u.s. national income and product accounts. Survey of current business, Bureau of Economic Analysis, Juillet 1997.

[60] Oded Galor. Convergence ? inferences from theoretical models. *The Economic Journal*, 106 :1056–1069, Juillet 1996.

[61] Guillaume Gaulier, Hurlin Christophe, et Jean-Pierre Philippe. Testing convergence : A panel data approach. *Annales d'Économie et de Statistique*, 55-56 :411–427, 1999.

[62] Alexander Gerschenkron. *Economic Backwardness in Historical Perspective*. University of Chicago Press, 1952.

[63] I. Gijbels, E. Mammen, B.U. Park, et Simar L. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, 1999.

[64] Jérôme Glachant et Charles Vellutini. Mobilité du capital et convergence des revenus dans le modèle néo-classique. mimeo, CREST, EPEE, Novembre 2000.

[65] Christian Gourieroux et Alain Monfort. *Séries Temporelles et Modèles Dynamiques*. Economica, 1995.

[66] Christian Gourieroux et Alain Monfort. *Statistique et moèles économétriques*, volume 1 - Notions générales, Estimation, Prévisions, Algorithmes. Economica, 1996.

[67] S. Grosskopf. Some remarks on productivity and its decompositions. mimeo, 2001.

[68] Wouter J. den Haan. Convergence in stochastic growth models : the importance of understanding why income levels differs. *Journal of Monetary Economics*, 35 :65–85, 1995.

[69] Robert Hall et Charles Jones. Why do some countries produce so much more output per worker than others ? *Quarterly Journal of Economics*, 1999.

[70] James D. Hamilton. *Time Series Analysis*. Princeton University Press, New Jersey, 1994.

[71] Bruce Hansen. Sample splitting and threshold estimation. *Econometrica*, 68 :575–603, 2000.

[72] Wolfang Härdle. *Applied Nonparametric Regression*, volume 19 of *Econometric Society Monographs*. Cambridge University Press, 1990.

[73] B. Hobijn, P.H. Franses, et M. Ooms. Generalizations of the kpss-test for stationarity. Working Paper 9802/A, Erasmus University of Rotterdam, 1998.

[74] Bart Hobijn et Philip Hans Franses. Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, 15(1), 2000.

[75] Charles R. Hulten. Technical change and the reproducibility of capital. *American Economic Review*, 65(5), Décembre 1975.

[76] Charles R. Hulten. Total factor productivity : a short biography. Working Paper 7471, NBER, Janvier 2000.

[77] Nazrul Islam. Growth empirics : A panel data approach. *Quarterly Journal of Economics*, 110(443) :1127–1170, Novembre 1995.

[78] Nazrul Islam. International comparison of total factor productivity : A review. mimeo, Department of Economics, Emory University, 2001.

[79] Henderson Daniel J. et Robert R. Russell. Human capital and convergence : A production frontier approach. Working paper, University of Caliofornia, Riverside, 2001.

[80] Peter Klenow et Andres Rodriguez-Clare. The neoclassical revival in growth economics : Has it gone too far ? *NBER Macroeconomics Annual*, pages 73–102, 1997.

[81] A. Kneip, B. U. Park, et L. Simar. A note on the convergence of nonparametric dea efficiency measures. *Econometric Theory*, 14 :783–793, 1998.

[82] Alois Kneip et Léopold Simar. A general framework fo frontier estimation with panel data. Discussion paper, CORE, Août 1995.

[83] Irving B. Kravis, Alan Heston, et Robert Summers. *World Product and Income, Internationnal comparisons of real gross product*. The Johns Hopkins University Press, 1982.

[84] Subodh Kumar et Robert R. Russell. Technological change, technological catch-up, and capital deepening : Relative contributions to growth and convergence. *The American Economic Review*, 92(3) :502–526, 2002.

[85] Subal C. Kumbhakar et Lovell C.A. Knox. *Stochastic Frontier Analysis.* Cambridge University Press, 2000.

[86] Daniel Kwiatkowski, Peter C.B. Phillips, Peter Schmidt, et Youngcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root ? *Journal of Econometrics*, 54 :159–178, 1992.

[87] Junsoo Lee. On the power of stationarity tests using optimal bandwidth estimates. *Economics Letters*, 51 :131–137, 1996.

[88] Andrew Levin et Chien-Fu Lin. Unit root tests in panel data : Asymptotic and finite samples properties. Working Paper 92-23, Department of Economics, University of California, San Diego, Mai 1992.

[89] Andrew Levin et Chien-Fu Lin. Unit root tests in panel data : New results. Working Paper 93-56, Department of Economics, University of California, San Diego, Décembre 1993.

[90] Ross Levine et David Renelt. A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82(2) :942–963, 1992.

[91] Q. Li. Nonparametric testing of closeness between two unknown distribution functions. *Econometric Reviews*, 15 :261–274, 1996.

[92] Zhenjuan Liu et Stengos Thanasis. Non-linearities in cross-country growth regressions : a semiparametric approach. *Journal of Applied Econometrics*, 14 :527–538, 1999.

[93] Robert E. Lucas. *Models of Business Cycles.* Basil Blackwell, 1987.

[94] G.S. Maddala et In-Moo Kim. *Unit Roots, Cointegration and Structural Change.* Cambridge, University Press, 1998.

[95] N.Gregory Mankiw, David Romer, et David N. Weil. A contribution to the empirics of economic growth. *Quaterly Journal of Economics*, 107 :407–437, 1992.

[96] Robin Marris. Comparing the incomes of nations : A critique of the international comparison project. *Journal of Economic literature*, XXII :40–57, 1984.

[97] Philippe Martin. Convergence et politiques régionales en europe. La lettre du CEPII 159, CEPII, Juillet 1997.

[98] Andreu Mas Colell, Michael D. Whinston, et Jerry R. Green. *Microeconomic Theory.* Oxford University Press, 1995.

[99] Kevin M. Murphy, Andrei Shleifer, et Vishny Robert W. Industrialization and the big push. *Quaterly Journal of Economics*, 106 :503–530, 1989.

[100] V. Nehru et A. Dhareshwar. A new database on physical capital stock : Sources, methodology and results. *Revista de Análisis Económico*, 8 :37–59, 1993.

[101] V. Nehru, E. Swanson, et Dubey A. A new database on human capital stock in developing and industrial countries : Sources, methodology and results. *Journal of Development Economics*, 46 :379–401, 1995.

[102] Richard R. Nelson et Edmund S. Phelps. Investments in humans, technological diffusion and economic growth. *The American Economic Review (American Economic Association : Papers and Proceedings)*, 61 :69–75, 1966.

[103] W K. Newey et Kenneth West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61 :631–653, 1994.

[104] Sam Ouliaris, Joon Y. Park, et Phillips Peter C.B. *Testing For a Unit Root in the Presence of a Maintained Trend*, pages 7–28. D. Reidel Publiher Company, Holland, 1989.

[105] Howard Pack. Endogenous growth theories : Intellectual appeal and empirical shortcomings. *Journal of Economic Perspective*, 8(1) :55–72, 1994.

[106] Adrian Pagan et Aman Ullah. *Nonparametric Econometrics*. Cambridge University Press, 1999.

[107] B. Park, L. Simar, et Weiner C. The FDH estimator for productivity efficiency scores : Asymptotic properties. *Econometric Theory*, 16 :855–877, 2000.

[108] P. Perron. The great crash, the oil-price shock, and the unit root hypothesis. *Econometrica*, 57 :1361–1401, 1989.

[109] Peter C.B. Phillips. Towards a unified asymptotic theory of autoregression. *Biometrika*, 74 :535–548, 1987.

[110] Peter C.B. Phillips. Spectral regression for cointegrated time series. In William Barnett, editor, *Nonparametric and Semiparametric Methods in Economics and Statistics*, chapter 16, pages 413–435. Cambridge University Press, 1991.

[111] Peter C.B. Phillips. New tools for understanding spurious regressions. *Econometrica*, 66(6) :1299–1325, 1998.

[112] Peter C.B. Phillips et Stephen N. Durlauf. Multiple time series regression with integrated processes. *Review of Economic Studies*, 53 :473–495, 1986.

[113] Peter C.B. Phillips et Hyungsik Moon. Estimation of autoregression roots near unity using panel data. Discussion paper, Yale University, Janvier 1999.

[114] Peter C.B. Phillips et Hyungsik Moon. Linear regression limit theory for nonstationary panel data. Discussion paper, Yale University, Avril 1999.

[115] Peter C.B. Phillips et Joon Y. Park. Statistical inference in regressions with integrated processes : Part 1. *Econometric Theory*, 4 :468–497, 1988.

[116] Peter C.B. Phillips et Donggyu Sul. Dynamic panel estimation and homogeneity testing under cross section dependence. Discussion paper, Yale University, Mai 2002.

[117] Dimitris N. Politis, Joseph P. Romano, et Michael Wolf. *Subsampling*. Springer-Verlag, New York, 1999.

[118] James Proudman, Stephen Redding, et Marco Bianchi. Is international openness associated with faster economic growth ? Working paper, Bank of England, 1997.

[119] Danny T. Quah. International patterns of growth : I. persistence in cross-countries disparities. Working paper, MIT, Mars 1990.

[120] Danny T. Quah. Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95 :427–443, 1993.

[121] Danny T. Quah. Twin peaks : Growth and convergence in models of distribution dynamics. *The Economic Journal*, 106 :1045–1055, Juillet 1996.

[122] Danny T. Quah. Empirics for growth and distribution : Polarization, stratification and convergence clubs. *Journal of Economic Growth*, 2(1) :27–59, Mars 1997.

[123] Danny T. Quah. Regional cohesion from local isolated auctions : II conditioning. Discussion Paper 379, CEP, Décembre 1997.

[124] Subhash C. Ray et Evangelia Desli. Productivity growth, technical progress and efficiency change in industrialized countries : Comment. *The American Economic Review*, 87 :1033–1039, 1997.

[125] Tyrell R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, USA, 1970.

[126] Paul Romer. The origins of endogenous growth. *Journal of Economic Perspective*, 8(1) :3–22, 1994.

[127] Xavier Sala-i-Martin. The classical approach to convergence analysis. *Economic Journal*, 106 :1019–1036, 1996.

[128] Xavier Sala-i-Martin. I just ran two million regressions. *American Economic Review, AEA Papers and Proceedings*, 87 :178–183, 1997.

[129] J.D. Sargan et Alok Bhargava. Maximum likelihood estimation of regression models with first moving average errors when the root lies on the unit circle. *Econometrica*, 51(3) :799–820, 1983.

[130] Peter Schmidt et Peter C.B. Phillips. Lm tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54 :257–287, 1992.

[131] Etsuro Shioji. Convergence in panel data : Evidence from the skipping estimation. Working paper, Department of Economics and Business, Universitat Pompeu Fabra, Août 1997.

[132] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*, volume 26 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York, 1986.

[133] Léopold Simar. Detecting outliers in frontier models : A simple approach. Discussion paper 146, Institut de statistique, Université Catholique de Louvain, 2001.

[134] Léopold Simar et Paul W. Wilson. Sensitivity analysis of efficiency scores : How to bootstrap in nonparametric frontier models. *Management Sciences*, 44(1) :49–61, 1997.

[135] Léopold Simar et Paul W. Wilson. Estimating and bootstraping malmquist indices. Discussion paper, Institut de statistique, Université Catholique de Louvain, Octobre 1998.

[136] Léopold Simar et Paul W. Wilson. A general methodology for bootstrapping in nonparametric frontier models. Discussion paper, Institut de statistique, Université Catholique de Louvain, Avril 1998.

[137] Léopold Simar et Paul W. Wilson. Nonparametric tests of returns to scale. Discussion paper, Institut de statistique, Université Catholique de Louvain, Juin 1998.

[138] Léopold Simar et Paul W. Wilson. Productivity growth in industrialized countries. Discussion paper, Institut de statistique, Université Catholique de Louvain, Février 1998.

[139] Léopold Simar et Paul W. Wilson. Statistical inference in nonparametric frontier models : The state of the art. Discussion paper, Institut de statistique, Université Catholique de Louvain, Mars 1999.

[140] Robert M. Solow. A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70 :65–94, 1956.

[141] Robert M. Solow. Technical change and the aggregate production function. *Review of Economics and Statistics*, 39 :312–320, 1957.

[142] Robert M. Solow. *Growth Theory : An Exposition*. Cambridge University Press, New-York, 1970.

[143] Miguel St Aubyn. Convergence across industrialized countries (1890-1989) : New results using time series methods. Working Paper 2/96, Universidade Técnica de Lisboa, 1996.

[144] James Stock. Unit roots, structural breaks and trends. In Engle et McFadden, editors, *Handbook of Econometrics, volume IV*, pages 2739–2841. Elsevier Science, 1994.

[145] James H. Stock. Confidence intervals for the largest autoregressive root in u.s. macroeconomic time series. *Journal of Monetary Economics*, 28 :435–459, 1991.

[146] Robert Summers et Alan Heston. Penn world tables (mark 5) : An expanded set of international comparisons, 1950-1988. *Quaterly Journal of Economics*, 2(106) :327–368, Mai 1991.

[147] Robert Summers et Alan Heston. International price and quantity comparisons : Potentials and pitfalls. *The American Economic Review (AEA, Papers and Proceedings)*, 86(2) :20–24, Mai 1996.

[148] Katsuko Tanaka. *Time Series Analysis, Nonstationary and Noninvertible Distribution Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1996.

[149] Johathan Temple. The new growth evidence. Working paper, Hertfort College, Oxford, Mars 1998.

[150] Ruey S. Tsay et George C. Tiao. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma model. *Journal of the American Statistical Association*, 79 :84–96, 1984.

[151] Jaume Ventura. Growth and interdependence. *Quarterly Journal of Economics*, pages 57–84, 1997.

[152] Kenneth D. West. A note on the power of least squares tests for a unit root. *Economic Letters*, 24 :249–252, 1987.

[153] Edward N. Wolff. Capital formation and productivity convergence over the long run. *American Economic Review*, 81(3) :565–579, 1991.

[154] A. Young. The tyranny of numbers : Confronting the statistical realities of the east asian growth experience. *Quarterly Journal of Economics*, 110(3) :641–680, 1995.

[155] Eric Zivot et Donald W.K. Andrews. Further evidence on the great crash, the oil-price shock, and the unit root hypothesis. *Journal of Business and Economic Statistics*, 10 :251–270, 1992.