

Analyse cliométrique des cycles de croissance de l'éducation en France (1815-2003) : Vers un modèle à seuil autorégressif

Jean-Pascal GUIRONNET¹

Version révisée de juillet 2005

***Résumé** : L'objectif de cet article est d'étudier l'évolution, sur longue période, des effectifs scolarisés français. Au cours du XIX^e et XX^e siècle, le développement du système éducatif s'est effectué non pas linéairement mais de manière convulsive. Partant de là, nous choisissons une formalisation économétrique en seuil afin d'identifier les ruptures de grandes ampleurs dans la composante cyclique de la série des effectifs scolarisés. Notre approche cliométrique met ainsi en évidence trois « cycles de croissance » de l'Éducation nationale et apporte une nouvelle interprétation économique de deux siècles d'histoire de la scolarisation.*

Mots clefs : Modèles à seuil ; éducation ; non linéarité.

JEL-classification : C22, I2.

¹ LAMETA, Université de Montpellier 1.
Adresse : LAMETA, Université de Montpellier 1, Faculté des sciences économiques, Espace Richter, Avenue de la Mer, E-mail : jpguironnet@lameta.univ-montp1.fr.

INTRODUCTION

L'objet de cette étude est d'identifier économétriquement les différentes phases de l'évolution des effectifs scolarisés français. En particulier, nous cherchons à savoir comment s'est effectué ce développement ? Quelles en furent les principales étapes et comment les interpréter ?

La majorité des travaux antérieurs, aussi bien des sociologues que des historiens, analyse l'évolution du système éducatif après la seconde guerre mondiale. Globalement, ces recherches distinguent deux explosions scolaires, l'une en 1960, l'autre à partir de 1985 (Prost, 1998). A la différence de ces études, notre contribution s'inscrit dans une approche cliométrique avec une analyse des effectifs scolarisés français sur deux siècles. Dans le prolongement des travaux de C. Diebolt (1999), nous poursuivons la même démarche d'histoire quantitative afin de mieux comprendre le développement de longue période du système éducatif. Si les analyses dans ce domaine consistent essentiellement à repérer et commenter les inflexions des effectifs scolarisés, les seuils déterminés (ou ruptures) dans cet article sont identifiés par une technique économétrique.

Plus précisément, les analyses des séries temporelles reposent généralement sur une modélisation ARIMA. Cependant, la mise en œuvre de ce type de processus nécessite d'effectuer implicitement l'hypothèse de linéarité qui constitue également une limite majeure puisqu'elle ne peut expliciter d'éventuelles asymétries dans l'évolution des variables. Par exemple, Neftçi (1984) trouve que le taux de chômage augmente plus vite qu'il ne diminue. Afin de s'amender du cadre restrictif imposé par l'hypothèse de linéarité, on utilise des modèles économétriques capables de gérer des dynamiques différentes selon la phase du cycle contrairement à un processus linéaire qui suppose une symétrie tout au long du cycle (Ben Salem et Perraudin, 2001). Dans cet article, nous présentons les modèles à seuil qui sont des cas particuliers des processus à non linéarités dans la moyenne. Les modèles de non linéarités en variance (ARCH) sont plus adaptés aux problématiques financières et diffèrent de notre thème de recherche (Guironnet, 2005).

Suivant cette approche empirique sur le long terme, nous mettons en évidence différents « cycles de croissance » du système éducatif et cherchons à en identifier les origines. La dimension historique de notre analyse, nous permet d'évaluer l'état des choses actuelles sur nos institutions scolaires. Ainsi, nous sommes confrontés aux grandes questions de la société en ce qui concerne : la légitimité de mesures volontaristes qui permettent à un nouveau public d'atteindre les degrés supérieurs du système éducatif et de la dévalorisation des diplômes.

Notre analyse est scindée en trois parties. Notre première section présente la spécification théorique des modèles à seuil et leur méthodologie d'estimation. Après avoir examiné notre série d'observations, nous détaillons le traitement économétrique. Nous terminons en dégagant une interprétation économique des différentes phases identifiées dans l'évolution de l'enseignement scolaire français.

D) LES MODELES A SEUIL

- *Spécification théorique*

Ce type de modélisation, défini par plusieurs régimes aux dynamiques distinctes, présente de nombreux intérêts. De tels modèles permettent de déterminer des processus différents dans les évolutions ascendantes ou descendantes d'une série économique. Afin d'expliquer ces asymétries, les modèles à seuil mettent en évidence les ruptures de grandes ampleurs dans une série temporelle. Cette approche s'inscrit ainsi comme un complément aux analyses quantitatives des évolutions de longue période qui peuvent parfois s'avérer complexes.

Parmi les modèles à seuil existants, nous retenons ceux dans lesquels la transition entre les régimes est brutale (modèle SETAR). Pour le cas des transitions douces, on fait référence aux modèles STAR (Granger, Teräsvirta, 1993 ; Teräsvirta, 1998 et Potter, 1999)². L'ensemble de ces modèles remet en cause la non stationnarité d'un processus établie dans un cadre linéaire. Par exemple, un processus peut suivre une marche aléatoire dans une zone centrale mais à contrario avoir un comportement stable dans les zones extrêmes (Caner et Hansen, 2001).

Les modèles SETAR (« *self-exciting threshold autoregressive* ») sont le prolongement des modèles de classe TAR (« *threshold autoregressive* ») introduits par Tong et Lim (1980). L'idée sous-jacente de ce type d'approche est qu'une relation peut être non linéaire sur la période globale considérée mais linéaire par sous périodes. La forme d'un modèle SETAR, pour un processus stochastique $\{x_t\}$, s'écrit :

$$x_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} x_{t-i} + \varepsilon_t^{(1)}, & \text{si } x_{t-d} < r_1 \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} x_{t-i} + \varepsilon_t^{(2)}, & \text{si } r_1 \leq x_{t-d} < r_2 \\ \dots\dots\dots \\ \phi_0^{(k)} + \sum_{i=1}^{p_k} \phi_i^{(k)} x_{t-i} + \varepsilon_t^{(k)}, & \text{si } x_{t-d} \geq r_{k-1} \end{cases} \quad (1)$$

Le processus SETAR, ci-dessus, est constitué de plusieurs morceaux de processus autorégressifs linéaires³. Le changement de processus dépend du paramètre de seuil r_j et de la valeur de x_{t-d} avec d le paramètre de délai (ou retard). Chaque équation représente un modèle AR linéaire d'ordre p_i où $i = 1, \dots, k$. Les erreurs $\{\varepsilon_t^{(1)}, \dots, \varepsilon_t^{(k)}\}$ sont des bruits blancs non corrélés. Dans le cas général (1), le modèle décrit par k régimes est noté $SETAR(k; p_1, \dots, p_k)$. Ce type de formalisation permet de considérer des structures de retards différentes pour chaque sous période et de mener ainsi une modélisation relativement souple.

² Outre les processus SETAR et STAR, il existe d'autres types de modèles à seuil. Pour plus de précisions, on peut consulter Guégan (1994) et pour un exposé des modèles à seuil en temps continu (processus CTAR), voir Brockwell et al. (1991).

³ Il est possible d'introduire des termes MA dans chaque régime, nous obtenons alors des processus SETARMA (Tong, 1983).

Lorsque, les ordres de retards sont identiques pour chaque équation, c'est-à-dire si $p_i = p \forall i$, les modèles SETAR se ramènent aux modèles TAR d'ordre p qui s'écrivent :

$$x_t = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^p \phi_i^{(1)} x_{t-i} + \varepsilon_t^{(1)}, & \text{si } x_{t-d} < r_1 \\ \phi_0^{(2)} + \sum_{i=1}^p \phi_i^{(2)} x_{t-i} + \varepsilon_t^{(2)}, & \text{si } r_1 \leq x_{t-d} < r_2 \\ \dots\dots\dots \\ \phi_0^{(k)} + \sum_{i=1}^p \phi_i^{(k)} x_{t-i} + \varepsilon_t^{(k)}, & \text{si } x_{t-d} \geq r_{k-1} \end{cases} \quad (2)$$

Les variables x_{t-i} sont des valeurs de l'endogène retardée, mais il est également possible d'introduire des variables explicatives autres que l'endogène retardée. Selon Tong (1990), Caner et Hansen (2001), la variable de seuil est une variable qui ne figure pas parmi les variables explicatives (endogène ou exogène). Cependant, puisque cette variable détermine les changements de régimes, il semble naturel de l'introduire parmi les variables explicatives. Le choix de la variable de transition (ou seuil) peut être inspiré par la théorie économique mais une solution athéorique est aussi envisageable : par exemple comme le précise le paragraphe suivant, la procédure de spécification du modèle à seuil permet de décider quelle variable retenir.

La prochaine section présente la procédure d'estimation d'un modèle SETAR qui s'inspire principalement de la méthode de Tsay (1989). Même si cette méthode est considérée comme intuitive, elle produit cependant des résultats fiables⁴.

- *Choix de la variable de transition et emplacement des seuils*

La variable de transition étant inconnue, le test de linéarité de Tsay (1989) s'effectue sur toutes les variables retardées possibles. Le choix de d est heuristique, il repose sur l'idée que si un modèle SETAR est nécessaire, il faut retenir le paramètre de délai pour lequel la linéarité est la plus fortement rejetée. Le paramètre de retard optimal est donc celui qui maximise la statistique $Q(m)$ de Tsay. Alternativement, la procédure d'identification de Tong permet aussi de déterminer la variable de seuil. Elle consiste à répéter l'analyse avec les différentes valeurs possibles de d et de retenir comme valeur optimale du paramètre de délai, celle qui minimise le critère AIC⁵ global sachant que :

$$AIC_{global} = \sum_{j=1}^k AIC_j(r_{k-1}^*), j = 1 \dots k. \quad (3)$$

La validité des critères d'information Akaike (1973) ou Schwarz (1978) dans un contexte de non linéarité peut être discutée (Kapetanios, 2001) ; néanmoins faute de mieux, ils sont toujours utilisés.

⁴ Hansen (1996) a construit une procédure alternative. Pour une comparaison des deux méthodes, voir Ben Salem et Perraudin (2001).

⁵ « Akaike Information Criterion » (Akaike, 1974).

La variable de transition déterminée, il reste à identifier l'emplacement des seuils de façon graphique. La méthode employée consiste à représenter graphiquement les t de Student des coefficients régressifs, ou les résidus, estimés de manière récursive en fonction de la variable de seuil ordonnée de façon croissante (*cf.* annexes). Dans ce cadre, le processus associé aux r premières observations est linéaire mais dès que l'on introduit l'observation dont la valeur correspond au seuil, le processus devient une combinaison de deux régressions. La trajectoire des résidus récursifs (et les t de Student des coefficients estimés, ...) va alors se déformer en passant par la valeur du seuil qui est ainsi localisé à la première rupture repérée (Mignon et Lardic, 2002). Dans la dernière étape, connaissant l'emplacement des seuils et le paramètre de délai, on estime le modèle SETAR par les méthodes de régression usuelles.

II) APPLICATION : L'EVOLUTION DES EFFECTIFS SCOLAIRES FRANÇAIS

La reconstitution de la série chronologique des effectifs scolaires du XIX^e et XX^e siècle repose sur plusieurs sources de données plus ou moins récentes. Notre série d'observations reprend la base de donnée élaborée par C. Diebolt (1999)⁶. Ces différentes séries de base ont été élaborées selon diverses méthodes statistiques, conduisant à effectuer certaines corrections pour harmoniser la série chronologique. Cependant, ces arrangements sont manifestement insuffisants : nous observons que la scolarisation des 5-14 ans atteint un maximum dans les années 1880 et plafonne par la suite (*cf.* figure 2). Il est donc probable que les observations du XIX^e siècle soient surestimées.

La croissance des effectifs scolaires comporte une tendance longue évidente, parcourant l'ensemble du XIX^e et XX^e siècle. Afin d'isoler les cycles de croissance, nous rejoignons le cadre d'analyse de Kapetanios (2003) en estimant la composante tendancielle. L'objectif est de déterminer le processus stochastique $\{Y_t\}$ avec la tendance $\{Y_t^\tau\}$. A cette fin, nous prenons les logarithmes de la série originelle :

$$y_t = 100 \log(Y_t) \text{ et } y_t^\tau = 100 \log(Y_t^\tau) \quad (4)$$

Pour estimer la tendance, diverses méthodes sont possibles : le filtre d'Hodrick-Prescott (1997), la décomposition de Beveridge et Nelson (1981)⁷... En supposant, que la série des effectifs scolaires français comporte une tendance déterministe⁸, nous utilisons le filtre d'Hodrick et Prescott (HP dans la suite).

Le choix du paramètre λ du filtre HP est généralement fixé à 100 pour des données annuelles et 1600 pour des données trimestrielles. La détermination de ce paramètre étant en partie arbitraire, conformément à la plupart des travaux sur séries annuelles, nous fixons ce paramètre égal à 100. Pourtant pour modéliser des ruptures, il peut paraître séduisant de choisir un paramètre supérieur à 100 puisque la tendance devient moins sensible à la conjoncture et la partie cyclique plus importante. Mais, une augmentation du paramètre λ n'a pas d'impact important⁹.

⁶ Complétée, pour les plus récentes observations, par les chiffres publiés dans la série «évaluation et statistiques».

⁷ Pour les différentes décompositions tendance-cycle, voir Doz et al. (1995).

⁸ Les tests de racine unitaire de Dickey-Fuller (1979) et Phillips-Perron (1988) n'indiquent aucune racine unitaire, la série est générée par un processus ARMA.

⁹ Nous avons aussi estimé un modèle SETAR avec un paramètre de filtrage λ fixé à 400. Dans ce cas, les ruptures repérées sont similaires à celles retenues dans l'article (différences de l'ordre de 2 ans).

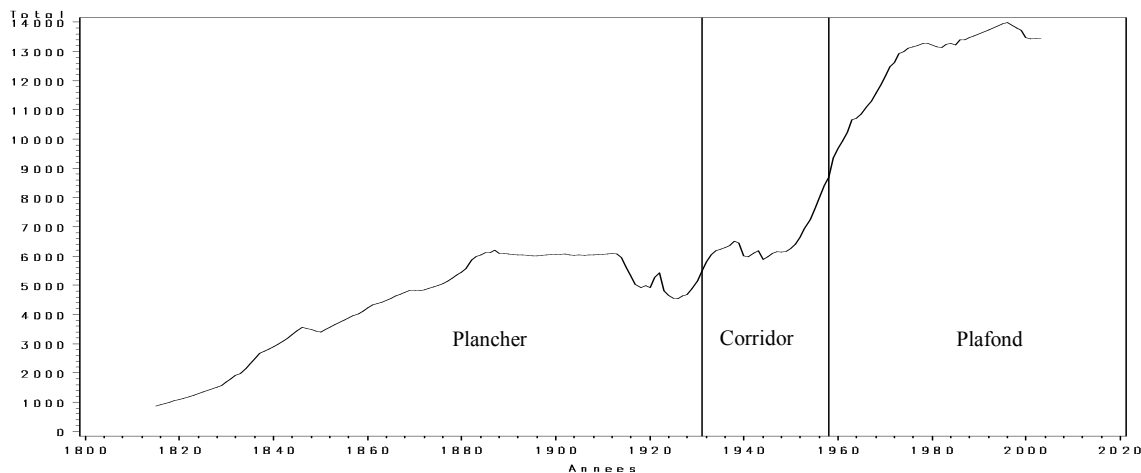
Nous obtenons après filtrage : $x_t = 100 \log(X_t) = 100 \log(Y_t / Y_t^t) = y_t - y_t^t$ où x_t représente l'écart en pourcentage de y_t par rapport à sa tendance. Notons que la série filtrée est stationnaire, le processus cyclique ne présente donc aucune racine unitaire. Selon les critères d'informations BIC¹⁰ et AIC, x_t serait un processus autorégressif d'ordre 2 sans constante. Cependant, si dans le modèle global cette dernière n'est pas significative, il est possible qu'elle le soit dans certaines sous périodes (Tsay, 1989). La constante pouvant mettre en évidence des ruptures, nous choisissons de l'inclure dans la recherche des seuils.

D'après les statistiques calculées du test de Tsay¹¹, l'hypothèse nulle de linéarité est la plus fortement rejetée¹² pour le paramètre de retard d égal 1. Suivant ce résultat, nous localisons à partir des graphiques de l'annexe 4 deux seuils $r_1 = -0.83$ et $r_2 = 0.55$. Nous identifions ainsi trois régimes ou « cycles de croissance » dans la courbe des effectifs scolarisés français, correspondant respectivement aux années 1931 et 1958 de la série non filtrée. L'estimation du modèle SETAR (3 ; 2, 1, 1) donne :

$$\begin{cases} x_t = -1.04x_{t-1} + 0.45x_{t-2} + \varepsilon_t & \text{si } -\infty < x_{t-1} \leq -0.83 \text{ (régime plancher)} \\ x_t = -0.75x_{t-1} + \varepsilon_t & \text{si } -0.83 < x_{t-1} \leq 0.55 \text{ (régime corridor)} \\ x_t = -0.73x_{t-1} + \varepsilon_t & \text{si } 0.55 < x_{t-1} \leq +\infty \text{ (régime plafond)} \end{cases} \quad (5)$$

Notons que la décomposition de la série chronologique en trois régimes a sensiblement amélioré l'estimation de la composante cyclique. Cependant, un découpage supplémentaire de la courbe des effectifs scolarisés aurait pu être envisageable. Toutefois, l'objectif de l'analyse est de dégager uniquement les principales phases de l'éducation nationale et donc de repérer seulement les plus grandes ruptures dans notre série d'observations.

Figure 1. Total (en millions) des effectifs scolarisés en France (1815-2003)



D'après la figure 1, l'évolution des effectifs scolarisés en France apparaît bien non linéaire mais par paliers ou en « marches d'escalier » à des rythmes variables et pour des raisons différentes (Chauvel, 1998). La natalité est l'un des facteurs expliquant le ralentissement ces

¹⁰ « Bayesian Information Criterion », (Schwartz, 1978).

¹¹ La statistique $Q(m)$ vaut respectivement pour $d=1$ et $d=2$: 26.09 et 14.84.

¹² Toutefois, nous avons estimé un modèle SETAR avec un paramètre de retard d égal 2. Confirmant les résultats du test de non linéarité, ce dernier modèle produit des résultats moins satisfaisants.

dernières années, elle reste élevée en France jusqu'en 1973 et participe implicitement à la croissance des flux scolaires. Depuis, elle régresse irrégulièrement mais sensiblement.

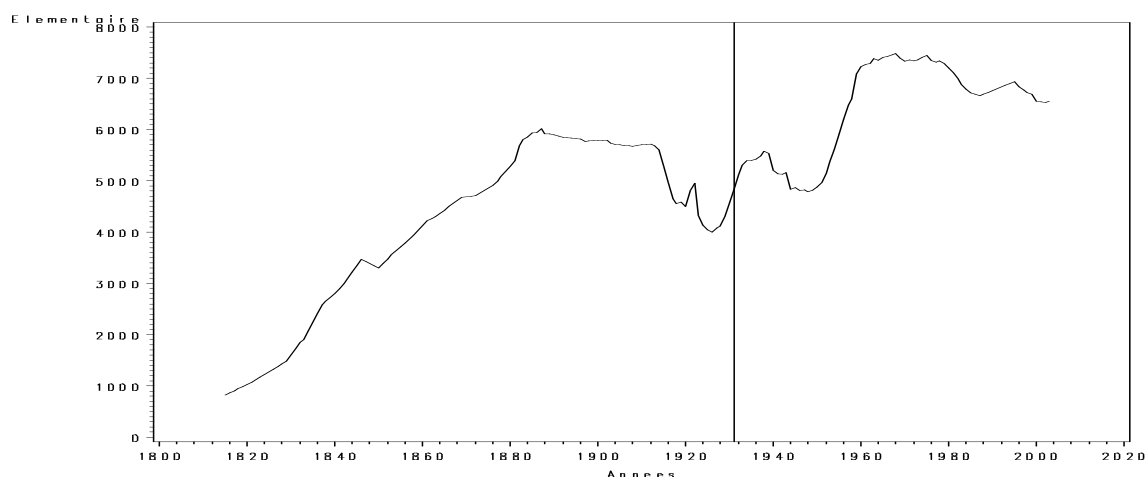
III) INTERPRETATIONS

L'évolution des effectifs scolarisés n'est, à l'évidence, pas continue mais bien à caractère convulsif. Le modèle SETAR révèle des discontinuités dans deux phases ascendantes de la série originelle, en 1931 et 1958. Les trois processus distincts sont nommés respectivement « plancher », « corridor » et « plafond » (cf. figure 1). La première période est synonyme d'une croissance lente, la seconde d'une accélération vigoureuse et le dernier régime est caractérisé par une croissance soutenue. Plus précisément, cette dernière phase regroupe les deux « cycles de croissance » habituellement retenus par les sociologues et les historiens (Chauvel, 1998 ; Prost, 1998). Suivant ces résultats, nous cherchons à identifier les raisons de ces ruptures dans la composante cyclique des effectifs scolarisés français.

- *Période de 1815-1931 : vers un grand « service public d'enseignement »*

Cette période est caractérisée essentiellement par l'évolution des effectifs du primaire¹³, les autres ordres¹⁴ éducatifs (secondaire et supérieur) comportant une proportion très faible d'effectifs scolarisés. Le cursus lycéen forme une élite restreinte : de 1815 à 1900, moins de 4% des 10-17 ans y sont scolarisés. « *Au 19^{ème} siècle, la statistique scolaire c'est d'abord, celle de l'enseignement primaire. Par sa précocité, son ampleur et sa régularité, elle dépasse les statistiques des autres ordres d'enseignement auxquelles elle sert parfois de modèle.* » (Luc, 1986). En 1931, les effectifs élémentaires restent encore largement majoritaires avec une proportion de plus de 88% des effectifs totaux, contre respectivement 10% et 1% pour le secondaire et le supérieur.

Figure 2. Total (en millions) des effectifs primaires en France (1815-2003)



La courbe des effectifs élémentaires, similaire à celle des effectifs totaux sur cette période, présente successivement une phase ascendante et stagnante (cf. figure 2). La première s'explique pour trois raisons : la loi Guizot de 1833, faisant obligation à toutes les communes de se doter d'une école primaire, qui ouvre la voie à la réalisation d'un grand « service public d'enseignement » ; le principe d'obligation scolaire des enfants de 6-13 ans en mars 1882

¹³ L'enseignement primaire comprend la scolarisation préélémentaire (enfants de moins de 6 ans) et les effectifs élémentaires (enfants de plus de 6 ans) dont la durée de formation est en moyenne de 5 ans.

¹⁴ L'école par ordres se transforme en une école par degrés (ou cycles) avec la réforme Berthoin (1959).

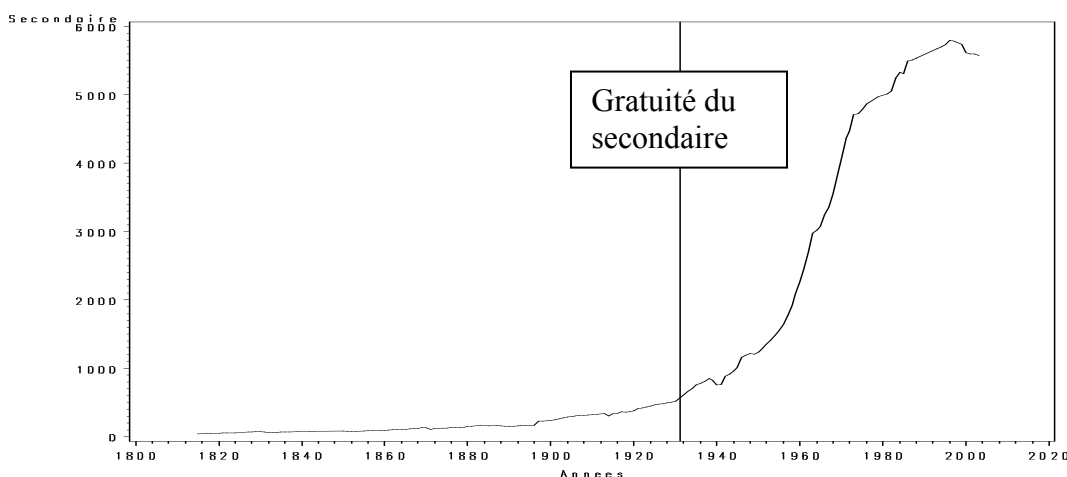
(sous la III République), et une conjoncture économique favorable. A contrario, la période de la partie stagnante est caractérisée par une croissance économique atone. Malgré les lois Ferry, gratuité du primaire et laïcité de l'éducation, l'effet défavorable de la situation économique reste dominant. La demande d'enseignement des ménages pour leurs enfants serait ainsi particulièrement sensible aux conjonctures économiques : un contexte économique défavorable, installant une incertitude chez les familles, se traduit par une diminution de la demande de formation et un ralentissement du développement des effectifs scolaires.

Plus généralement, la lente croissance du régime « plancher » s'explique aussi par de fortes réticences à l'égard de l'instruction du peuple. Selon un arrêté explicitant la loi sur l'obligation scolaire, « l'école primaire doit [...] préparer et prédisposer, en quelque sorte, les garçons aux futurs travaux de l'ouvrier et du soldat, les filles aux soins du ménage et aux ouvrages de femmes »¹⁵. A cette époque, le système éducatif était cloisonné et les divisions scolaires correspondant toujours aux divisions sociales, c'était au fils d'occuper la position sociale du père. La question de l'égalité des chances, qui sous tend la notion de méritocratie, n'émerge qu'après la Première Guerre mondiale ; cette prise de conscience sera le détonateur d'un nouveau régime que nous appellerons la première explosion scolaire.

- *Période de 1931-1958 : vers une massification du secondaire*

Les années 1930 s'ouvrent sur un nouveau cycle du système éducatif marqué par l'entrée progressive des jeunes dans le secondaire (cf. figure 3), alors que leurs aînés n'avaient connu que l'école élémentaire. Au cours de cette période, les effectifs du secondaire sont suffisamment importants pour influencer significativement l'évolution des effectifs totaux et ils se combinent désormais au processus des effectifs du primaire.

Figure 3. Total (en millions) des effectifs du secondaire en France (1815-2003)



Cet ordre d'enseignement, intégré dans la sphère publique par Napoléon, restera prépondérant jusqu'à la seconde guerre mondiale. Le recrutement de l'enseignement secondaire s'est élargi avec l'instauration de la gratuité (années 1930), l'absorption des écoles primaires supérieures (1943), l'accueil en lycée des élèves issus des cours complémentaires (années 1950). Par ailleurs, l'exigence de l'unification administrative du système éducatif rebondit après 1945 avec le plan Langevin-Wallon. Cette proposition suggérait l'idée d'une planification de l'extension progressive de l'école obligatoire jusqu'à l'âge de 18 ans, elle aurait fait preuve d'une clairvoyance historique si elle avait été mise en oeuvre.

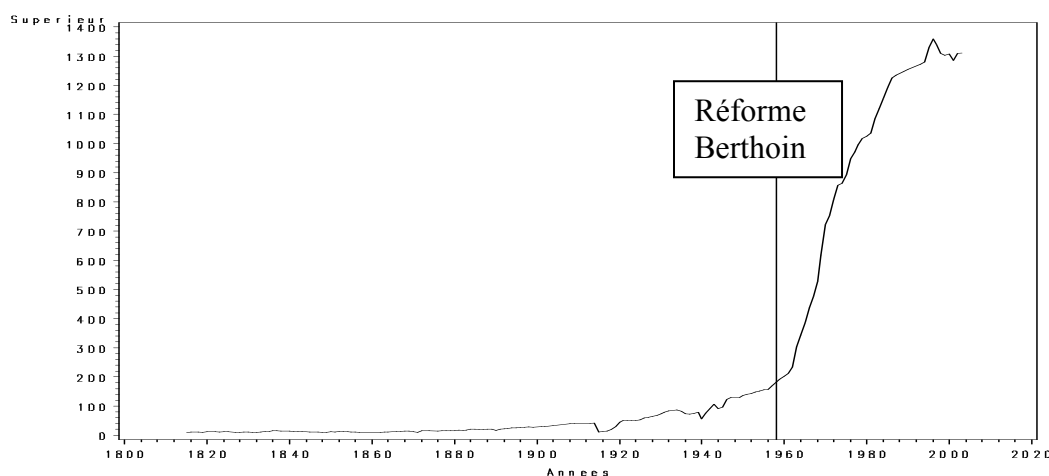
¹⁵ Cité par C. Lelièvre (1990).

Abstraction faite de la seconde guerre mondiale, ce régime connaît une très forte croissance du secondaire, sa part dans les effectifs totaux double et passe à plus de 22%, alors que les effectifs du supérieur restent toujours très faibles (2% des effectifs totaux). Ce second processus, ou première explosion scolaire, se caractérise donc principalement par la massification des effectifs du secondaire. Cependant, vers la fin des années cinquante¹⁶, l'évolution du système éducatif connaît un tournant décisif : l'afflux des élèves dans le supérieur.

- *Période de 1958-2003 : le décollage (« take off ») du supérieur*

Le seuil de 1958 représente l'origine du développement massif de l'enseignement supérieur. La réforme Berthoin (1959), adoptée sous la pression des besoins du développement économique et social¹⁷, prolonge l'obligation scolaire jusqu'à 16 ans et instaure un tronc commun au collège (6^{ième} et 5^{ième}). Le secondaire continu sa forte croissance jusqu'en 1990 et impulse un fort déversement d'étudiants dans le supérieur.

Figure 4. Total (en millions) des effectifs supérieur en France (1815-2003)



Cette ouverture du secondaire, dans les années 1960, a largement participé à une prolongation générale des cursus et favorisé l'accès de nouvelles catégories sociales aux études les plus longues. L'enseignement supérieur apparaît dès lors comme le grand bénéficiaire (cf. figure 4) de ce que l'on peut qualifier, de seconde explosion scolaire. L'idée que la formation est la solution à tous les problèmes économiques et sociaux tend à s'imposer, exemple le mot d'ordre lancé par Jean-Pierre Chevènement en 1985 : « *plus d'étudiants et de bacheliers [...], à terme 80% d'une classe d'âge doit atteindre le niveau du bac : c'est l'objectif que nous nous sommes fixés* ». La croissance des Trente glorieuses et la tertiarisation de l'emploi participent aussi au gonflement des effectifs lycéens ; la part du secondaire représente, aujourd'hui, près de 40% des effectifs scolarisés. Cette vague atteint d'emblée l'enseignement supérieur, leur proportion (hors IUT et IUFM) atteint plus de 15% des effectifs scolarisés en 2003. Cet important allongement de la durée des études est une sorte de révolution culturelle dans laquelle le diplôme occupe désormais une place privilégiée, « *la demande de titres scolaires de la part des étudiants se substitue de plus en plus à la demande d'éducation* » (Lévy-Garboua, 1976).

¹⁶ L. Chauvel (1998), en utilisant comme indicateur l'âge médian de fin d'études initiales par années de naissance, identifie une rupture en 1955.

¹⁷ Rapport officiel Rueff-Armand (1959) : « l'industrie devra créer, dans les douze années à venir, près de deux fois plus d'emplois qu'elle n'en a créé depuis le début du siècle ».

La croissance des effectifs scolaires s'explique aussi par les générations du *baby-boom* et l'essor de la scolarité des filles au cours des dernières décennies. A partir de 1958, les effectifs féminins du supérieur effectuent un rattrapage rapide sur leurs homologues masculins (cf. figure 6 en annexes). Cependant, la décision d'élever si fortement le niveau moyen de scolarisation en un temps si bref, était-elle légitime ?

- *Modification du marché du travail : vers un phénomène de suréducation*

La démocratisation de l'accès au secondaire a permis, en trois décennies, la massification de l'enseignement supérieur. Le système universitaire devient dès lors l'une des principales préoccupations des recherches en économie de l'éducation. Cependant, si le décloisonnement du système éducatif ouvre l'accès aux études longues au plus grand nombre, de nombreuses analyses sociologiques montrent que la fonction de tri social de l'école reste inchangée¹⁸ et la sélection simplement différée (Terrail, 1997). Maintenir plus longtemps l'égalité des chances a surtout conduit le système éducatif à repousser la première véritable différenciation des étudiants après la fin de la scolarité obligatoire et, de la reporter de plus en plus, au sein de l'enseignement supérieur (Béduwé et Planas, 2004).

L'actuel cycle avec une croissance sans précédent, ajouté à la crise d'emploi de notre économie, a transformé le fonctionnement du marché du travail : l'observation statistique permet sans conteste de conclure que les générations de plus en plus formées rencontrent des difficultés croissantes pour obtenir un emploi correspondant à leur niveau d'études. L'élévation des niveaux de formation satisfait des besoins réels mais au prix d'un phénomène massif de déclassement (« *overeducation* ») des jeunes diplômés (Forgeot, Gautié, 1997 ; Nauze-Fichet, Tomazini, 2002). L'important allongement général de la durée des études a ainsi produit, ces dernières décennies, une dégradation des perspectives d'accès aux postes élevés de la hiérarchie professionnelle, en particulier lors des premiers recrutements : pour obtenir un emploi, il faut désormais deux ou trois années d'études supplémentaires que les qualifications « normalement » requises. Cette forte élévation du niveau scolaire au cours de ce siècle suscite deux réactions contradictoires : un point de vue optimiste avec toujours de plus en plus de connaissances, l'autre pessimiste avec une dévalorisation des diplômes. Si au cours de ces dernières années, le déclassement des jeunes diplômés s'est largement amplifié ; l'évolution de ce phénomène ne s'est pas toujours développée de manière croissante.

Au cours de la « décennie dorée » (1965-1975), le marché de l'emploi était particulièrement favorable alors que pour les cohortes précédentes, malgré des conditions économiques toujours avantageuses, l'accès aux fonctions élevées de la hiérarchie professionnelle était plus restreint (Chauvel, 1998). Le phénomène de suréducation coïnciderait ainsi avec la fin de la première explosion scolaire et ne serait donc pas uniquement imputable à la conjoncture économique. Ce constat historique, ajouté à l'importante vague de déclassement observée de nos jours, nous amène à supposer une correspondance entre les périodes de dévalorisation des diplômés et les phases successives d'engorgement des effectifs d'un ordre éducatif (secondaire puis supérieur). Au cours de ces périodes, la multiplication des plus hautes certifications obtenues introduirait un scepticisme chez les employeurs sur la productivité potentielle du diplômé qui se manifeste par une surévaluation du niveau demandé pour l'emploi. Pour rétablir une certaine crédibilité dans le filtrage éducatif¹⁹, et sans occulter les considérations normatives (démocratisation de l'enseignement), les politiques éducatives facilitent par législation²⁰ les conditions d'accès du

¹⁸ Euriat et Thélot (1995) : l'inégalité des chances d'avoir un bac ne s'est que très légèrement atténuée.

¹⁹ L'université comme un filtre (Arrow, 1973).

²⁰ Avec, par exemple, un critère de sélectivité assoupli.

niveau supérieur lorsqu'un niveau inférieur se retrouve saturé²¹. Le système éducatif retrouve une sélectivité, certes différée, et d'un même pas la confiance des agents économiques. Le phénomène de suréducation s'amplifierait ainsi en fin de cycle de croissance et s'atténuerait au commencement d'une nouvelle phase du système éducatif.

Cependant, cette hypothèse reste difficilement testable pour deux raisons. Premièrement, nous manquons encore de recul historique pour établir véritablement cette correspondance. Par ailleurs, la fragilité du parallèle, entre les phases successives de saturation du système éducatif et le phénomène de suréducation, est qu'il repose sur un simple constat de simultanéité. Toutefois, comme le souligne Durkheim (1937), il constitue la base de la recherche de causalité : *« nous n'avons qu'un moyen de démontrer qu'un phénomène est cause d'une autre, c'est de comparer les cas où ils sont simultanément présents et absents et de chercher si les variations qu'ils présentent dans ces différentes circonstances témoignent que l'un dépend de l'autre »*.

Si le risque de suréducation est lié aux cycles de croissance du système éducatif, en alternant d'importante phase d'expansion et de récession, notre période contemporaine reste plus préoccupante pour deux raisons. Premièrement, il n'existe pas d'institution éducative « au-dessus » du système universitaire ; les politiques éducatives ne peuvent faciliter les conditions d'accès d'un ordre supérieur. Secondement d'après le modèle SETAR, le rythme de croissance des effectifs scolaires ces quatre dernières décennies est sans précédent par rapport aux évolutions antérieures (primaire et secondaire).

²¹ Lorsque le secondaire est saturé, on facilite l'accès au supérieur.

CONCLUSION

Par l'intermédiaire des modèles à seuil, nous avons identifié trois régimes distincts dans la série des effectifs scolaires français : un régime « plancher » avec une évolution lente contrairement au régime « corridor » et « plafond », où le rythme de croissance est élevé. Notre démarche d'histoire quantitative, nous permet de révéler deux explosions scolaires, l'une dans les années 1930 l'autre à partir de 1958, qui diffèrent de ceux retenus par les précédentes études. Au cours de chacun des régimes, nous remarquons une importance significative de chaque ordre éducatif dans la composition de la série chronologique. D'après nos résultats économétriques et un réexamen des études sociologiques, nous avons apporté une nouvelle interprétation de deux siècles d'histoire de scolarisation.

Historiquement, l'allongement de la durée des études a permis une ouverture considérable d'un accès au patrimoine scientifique et culturel de l'humanité, réservé autrefois à une élite restreinte d'individus. Mais si le système éducatif c'est « démographisé », il ne s'est pas pour autant démocratisé (A. Léger, 1991). En réalité, selon Bourdieu (1979), cette transformation n'a eu pour conséquence qu'une simple translation vers le haut de la structure de l'inégalité des chances²². De tels changements de structures ont sensiblement modifié la perception du diplôme avec d'importants mouvements successifs de déclassements. En effectuant le parallèle entre les vagues de suréducation et la massification en paliers des effectifs scolaires, nous avons approfondi le lien entre cycle de croissance du système éducatif et le phénomène de déclassement. Ainsi, nous observons que les deux phases de détérioration des perspectives d'accès aux postes élevés de la hiérarchie professionnelle correspondent chacune à une phase terminale d'un cycle. Suivant notre approche cliométrique, le sentiment de déclassement serait ainsi un phénomène rythmé par les cycles de croissance du système éducatif, en alternant des phases d'expansions brutales et des périodes de fortes récessions.

Aujourd'hui, l'accès à l'enseignement s'est fortement développé pour chaque ordre éducatif et une importante incertitude subsiste concernant les futures évolutions du système scolaire d'autant plus que le phénomène de suréducation est d'une ampleur sans précédent. Doit-on passer par un ré-haussement de la sélectivité dans les universités pour conserver une certaine crédibilité des certifications du supérieur ? La commission Fauroux (1996) montre que la tentation d'une sortie de la crise scolaire *par le bas*, d'un retour en arrière démolissant les politiques antérieures de massification scolaire, existe bel et bien au sein des sphères dirigeantes. On imagine cependant facilement l'ampleur des résistances²³ que sa mise en œuvre susciterait. Autre solution, vers laquelle on semble se diriger, celle d'une sortie de crise *par le haut* : la sélectivité doit-elle se faire dans le cadre d'institutions éducatives européennes ? Quoiqu'il en soit, des réformes profondes de l'enseignement semblent inévitables. Ces modifications structurelles seront certainement la marque d'un cycle historique achevé et d'une nouvelle étape dans l'évolution du système éducatif français.

²² La contribution de l'école à la reproduction des rapports de classes apparaît ainsi marquée d'invariance structurale (J-C. Passeron, 1991).

²³ Gamel (2000) : « Cette fuite devant les diplômes peut mettre du temps à apparaître, car elle signifie un bouleversement complet des repères traditionnels de l'insertion professionnelle, que les intéressés et surtout leurs parents, moulés dans leur jeunesse par le culte du diplôme associé à l'idéal républicain du mérite, ne semblent pas prêts à admettre aisément [...] ».

BIBLIOGRAPHIE

- AKAIKE, H. [1974], « A new look at the statistical model identification », *IEEE Transactions on Automatic and Control*, Series AC, 19, p.716-723.
- ARROW, K.J. [1973], « Higher education as a filter », *Journal of Public Economics*, 2.
- BEDUWÉ, C. ET PLANAS, J. [2004], « The effects of educational expansion on the functioning of the labour market: report of a comparative study », *Compare*, 34(1).
- AKAIKE, H. [1979], « A bayesian extension of the minimum AIC procedure », *Biometrika*, 66.
- BEN SALEM, M. ET PERRAUDIN, C. [2001], « Tests de linéarité, spécification et estimation de modèles à seuil : Une analyse comparée des méthodes de Tsay et de Hansen », *Economie et prévision*, 148, p.157-176.
- BEVERIDGE, S. ET NELSON, C. [1981], « A new approach to decomposition of economic time series into a permanent and transitory components with particular attention to measurement of the business cycle », *Journal of Monetary Economics*, 7, p.151-174.
- BOURDIEU, P. [1979], « La distinction, critique sociale du jugement », *Editions de Minuit*, Paris.
- BROCKWELL, P.J., HYNDMANN, R.J. ET GRUMWALD, G.K. [1991], « Continuous time threshold autoregressive models », *Statistica Sinica*, 1, p.401-410.
- CANER, M. ET HANSEN, B. [2001], « Threshold autoregression with a unit root », *Econometrica*, 69(6), p.1555-1596.
- CHAUVEL, L. [1998], « La seconde explosion scolaire : diffusion des diplômes, structure sociale et valeur des titres », *Revue de l'OFCE*, 66, p.5-36.
- DANTHINE, J.P. ET GIRARDIN, M. [1989], « Business cycle in Switzerland. A comparative study », *European Economic Review*, 33, p.31-50.
- DICKEY, D.A. ET FULLER, W.A. [1979], « Distribution of the estimation for autoregressive time series with a unit root », *Journal of The American Statistical Association*, 74, p.427-431.
- DIEBOLT, C. [1999], « Les effectifs scolarisés en France : XIXème et XXème siècles », *International Review of Education*, 45(2), p.197-213.
- DURKHEIM, E. [1981], « Les règles de la méthode sociologique », *PUF, nouv. Ed*, Paris.
- DOZ, C. RABAULT, G. ET SOBCZAK, N. [1995], « Décomposition tendance-cycle : Estimation par des méthodes statistiques univariées », *Economie et Prévision*, 120, p.73-93.
- ERTEL, J.E., ET FOWLKES, E.B. [1976], « Some algorithms for linear spline and Piecewise multiple linear regression », *Journal of the American Statistical Association*, 71, p.640-648.
- EURIAT, M. ET THELOT, C. [1995], « Le recrutement de l'élite scolaire depuis quarante ans », *Education et Formations*, 41.
- FORGEOT, G. ET GAUTIE, J. [1997], « Insertion professionnelle des jeunes et processus de déclassement », *Economie et Statistique*, 304-305, p.53-74.
- GAMEL, C. [1998], « Le diplôme, un signal en voie de dépréciation ? Le modèle de Spence réexaminé », *Revue d'économie politique*, 110(1), p.53-84.
- GRANGER, C.W.J. ET TERÄSVIRTA T. [1993], « Modelling nonlinear economic relationships », *Oxford University Press*, Oxford, 198p.
- GUEGAN, D. [1994], « Séries chronologiques non linéaires à temps discret », *Economica*.
- GUIRONNET, J.P. [2005], « Tests and modelling thresholds for time series », *Historical Social Research*, 30(2), p.329-336.
- HANSEN, B.E. [1996], « Inference when a nuisance parameter is not identified under the null hypothesis », *Econometrica*, 64(2), p.413-430.
- HODRICK, R. ET PRESCOTT, E. [1997], « Post-war U.S. business cycles: An empirical investigation », *Journal of Money, Credit and Banking*, 29[1], p.1-16.
- KAPETANIOS, G. [2001], « Model selection in threshold models », *Journal of Time Series*, 22(6), p.733-754.

- KAPETANIOS, G. [2003], « Threshold models for trended time series », *Empirical Economics*, 28(4), p.687-707.
- LEGER, A. [1991], « L'école et la réussite de tous », dans *Révolution par l'intelligence*, Editions Messidor, Paris, p.79-90.
- LELIEVRE, C. [1990], « Histoire des institutions scolaires (1789/1989) », *Nathan*, Paris.
- LEVY-GARBOUA, L. [1976], « Les demandes de l'étudiant ou les contradictions de l'université de masse », *Revue française de sociologie*, 17, p.53-80.
- LUC, J.-N. [1986], « L'illusion statistique », *Annales E.S.C.*, 41(4), p.887-911.
- MIGNON, V. ET LARDIC, S. [2002], « Econométrie des séries temporelles macroéconomiques et financières », *Economica*.
- NAUZE-FICHET, E. ET TOMASINI, M. [2002], « Diplôme et insertion sur le marché du travail : Approches socioprofessionnelles et salariale du déclassement », *Economie et Statistique*, 354, p.21-48.
- NEFTÇI, S.N. [1984], « Are economic time series asymmetric over the business cycle? », *Journal of Political Economy*, 92(2), p.307-328.
- PASSERON J.C. [1991], « Le raisonnement sociologique, l'espace non popérien du raisonnement naturel », *Nathan*, Paris.
- PHILLIPS, P.C. ET PERRON, P. [1988], « Testing for a unit root in time series regression », *Biometrika*, 75, p.335-346.
- POTTER, S.M. [1999], « Nonlinear time series modelling: An introduction », *Journal of Economic Surveys*, 13, p.505-528.
- PROST, A. [1998], « L'Education nationale depuis la Libération », *Cahiers français*, 285 [*Le système éducatif*], p.3-11.
- SCHWARZ, G. [1978], « Estimating the dimension of a model », *Annals of Statistics*, 6, p.461-464.
- TERÄSVIRTA, T. [1998], « Modelling economic relationships with smooth transition regressions », *Handbook of Applied Economic Statistics*, p.507-552.
- TERRAIL, J-P. [1997] [sous dir de], « La scolarisation de la France, critique de l'état des lieux », *La Dispute*, Paris.
- TONG, H. [1983], « Threshold models in nonlinear time series analysis », *Lectures Notes in Statistics*, 21, Springer Verlag.
- TONG, H. [1990], « Nonlinear time series: A dynamical system approach », *Oxford University Press*, Oxford.
- TONG, H. ET LIM, K.S. [1980], « Threshold autoregression, limit cycles, and cyclical data », *Journal of the Royal Statistical Society*, B42, p.245-292.
- TSAY, R.S. [1989], « Testing and modelling threshold autoregressive processes », *Journal of the American Statistical Association*, 84(405), p.231-240.

ANNEXES

Test de non linéarité de Tsay (1989):

Pour chaque variable de transition possible, $x_{t-d} = x_{t-i}$, les observations sont ordonnées selon les valeurs croissantes de la variable de seuil. Le modèle obtenu, avec un seuil entre la $r^{ième}$ et la $(r+1)^{ième}$ observation ordonnée, s'écrit :

$$x_{(c)} = \begin{cases} \phi_0^{(1)} + \phi^{(1)}x_{(c)} + \varepsilon_{(c)}^{(1)} & \text{pour les } r \text{ premières observations de } x_{t-d} \\ \phi_0^{(2)} + \phi^{(2)}x_{(c)} + \varepsilon_{(c)}^{(2)} & \text{pour les observations suivantes} \end{cases} \quad (\text{A1})$$

où c dénote le classement des observations.

A partir du modèle ordonné, on estime les coefficients par les moindres carrés récursifs²⁴ qui consiste à appliquer, pour une série quelconque Y_t , l'algorithme suivant :

$$\begin{aligned} \hat{\beta}_{m+1} &= \hat{\beta}_m + K_{m+1}[Y_{m+1} - x'_{m+1}\hat{\beta}_m], \\ D_{m+1} &= 1 + x'_{m+1}P_m x_{m+1}, \\ K_{m+1} &= P_m x_{m+1} / D_{m+1}, \\ &\text{et} \\ P_{m+1} &= (I - P_m \frac{x_{m+1}x'_{m+1}}{D_{m+1}})P_m \end{aligned} \quad (\text{A2})$$

avec $\hat{\beta}_m$, les estimations par les moindres carrés ordinaires pour les m premières données et P_m la matrice inverse de $X'X$ associée. x_{m+1} et Y_{m+1} représentent le vecteur des régresseurs et la variable expliquée de la prochaine observation respectivement.

Les résidus (l'équation A4 donne les résidus normés) sont calculés par :

$$\hat{a}_{m+1} = Y_{m+1} - x'_{m+1}\hat{\beta}_m \quad (\text{A3})$$

$$\hat{e}_{m+1} = \hat{a}_{m+1} / \sqrt{D_{m+1}}. \quad (\text{A4})$$

Une manière de tester la non linéarité en seuil est de régresser les résidus normés selon les observations ordonnées (A1) sur les régresseurs Y_{m+1} :

$$\hat{e}_{m+1} = \omega_0 + \sum_v^p w_v Y_{m+1} + \varepsilon_{m+1} \quad (\text{A5})$$

L'hypothèse nulle de linéarité s'écrit : $H_0 : \phi_i^{(1)} = \phi_i^{(2)}$ pour $i = 0, \dots, p$. La statistique correspondant à ce test est :

$$Q(m) = \frac{(\sum \hat{e}_i^2 - \sum \hat{\varepsilon}_i^2) / (p+1)}{\sum \hat{\varepsilon}_i^2 / (T - b - 2p - 1)} \quad (\text{A6})$$

²⁴ Ertel et Fowlkes (1976).

où b est le nombre d'observations utilisés pour initialiser les estimations récursives et T le nombre totale d'observations. $\hat{\varepsilon}_t^2$ sont les résidus ε_t standardisés et $\hat{\varepsilon}_t$ les résidus de (A5). Sous l'hypothèse nulle de linéarité, la statistique $Q(m)$ suit un Fisher à $(m+1)$ et $(T-b-2p-1)$ degrés de liberté. Suivant Tsay, on peut retenir $b = (T/10) + p$.

Le filtre de Hodrick-Prescott :

Ce filtre repose sur l'idée qu'une tendance doit être suffisamment lisse pour ne pas suivre toutes les inflexions de la série²⁵, sans pour autant s'écarter trop des mouvements de la série initiale. Soit une série temporelle Y_t qui est la somme d'une tendance T_t et d'une composante cyclique C_t : $Y_t = T_t + C_t$ pour $t = 1, \dots, n$.

La mesure de la trajectoire de T_t est la somme des carrés de ses différences secondes. Les C_t représentent les déviations par rapport à T_t (c'est-à-dire, $C_t = Y_t - T_t$). Pour déterminer les deux composantes de la série, il faut donc minimiser le programme suivant :

$$\sum_1^n C_t^2 + \lambda \sum_3^n (\Delta T_t - \Delta T_{t-1})^2 \quad (\text{A7})$$

Le second terme de (A7) mesure le degré de variabilité de la tendance, c'est-à-dire la somme des carrés des variations du taux de croissance de la tendance. Le paramètre λ représente le poids accordé aux écarts par rapport aux mouvements de la série initiale, il joue évidemment un rôle crucial dans la décomposition. Plus λ est élevé, plus l'objectif de lissage de la tendance est privilégié au détriment de l'adéquation à la série brute et donc plus la partie cyclique est importante.

Sous forme matricielle, (A7) s'écrit :

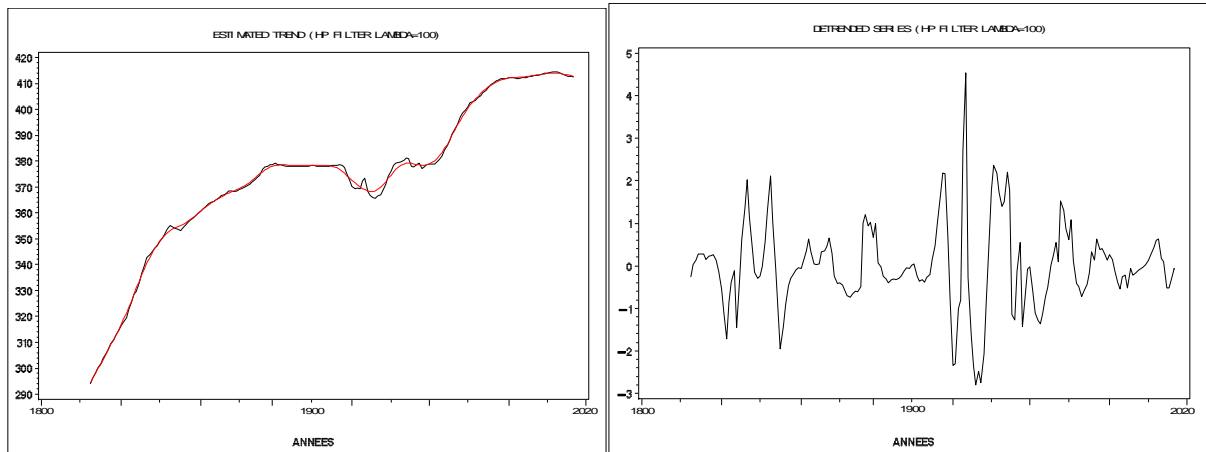
$$C_t' C_t + \lambda (KT_t)' KT_t \text{ où } KT_t = \begin{bmatrix} 1-2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 \\ \dots & & & & & & & \\ 0 & 0 & \dots & & 1-2 & 1 & & \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ \dots \\ T_n \end{bmatrix} \quad (\text{A8})$$

La solution²⁶ de (A8) est donnée par $T_t = A^{-1}Y_t$, où $A = I + \lambda K'K$.

²⁵ Implicitement attribuées au cycle

²⁶ Pour le détail de la démonstration, on peut consulter Danthine et Girardin (1989).

Figure 5. Filtre HP sur le logarithme de la série des effectifs scolarisés français



Grilles de recherche d'emplacement des seuils :

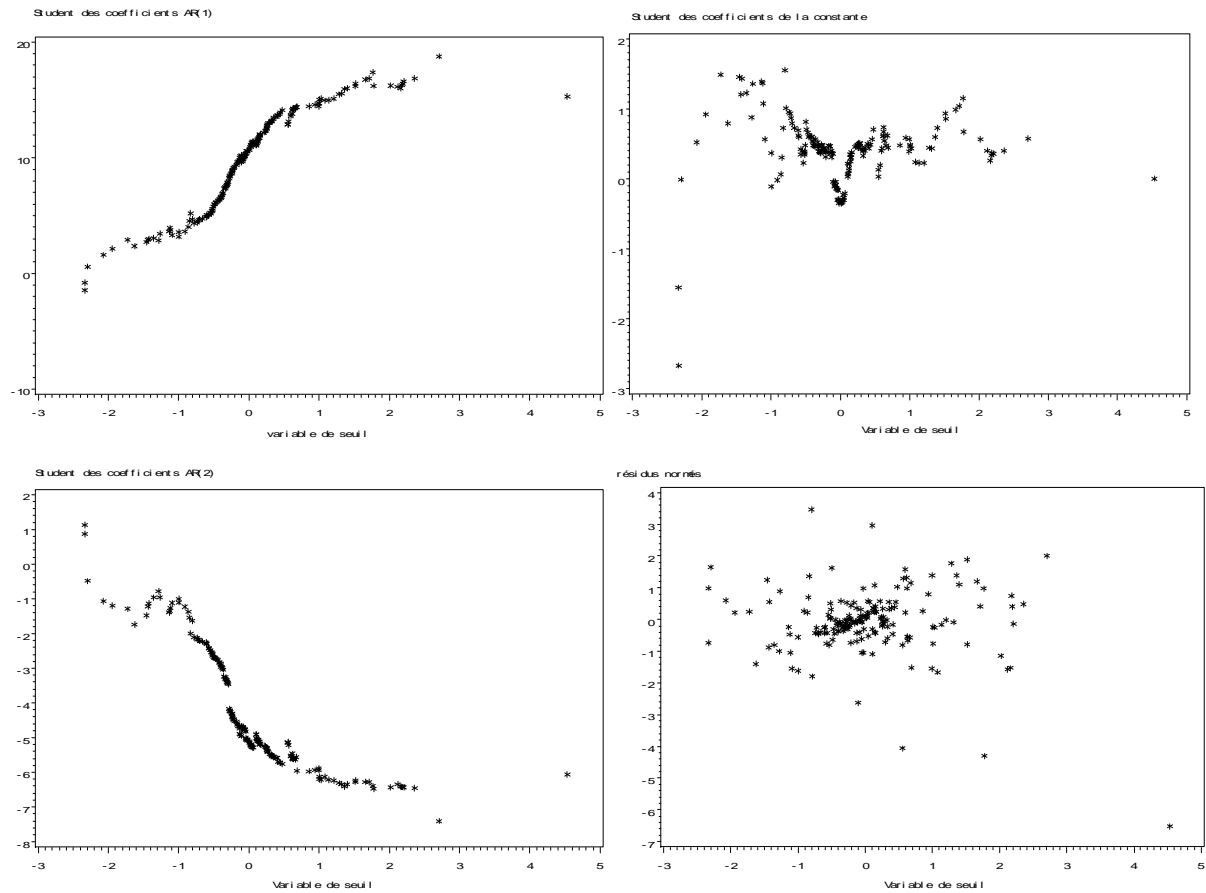


Figure 6. Total (en millions) des effectifs hommes/femmes du supérieur (1899-1963)

