

Tests de ruptures : une application au PIB tendanciel français

Hervé Le Bihan

Citer ce document / Cite this document :

Le Bihan Hervé. Tests de ruptures : une application au PIB tendanciel français. In: Économie & prévision, n°163, 2004-2. Élargissement de l'Union Européenne. pp. 133-154;

doi : 10.3406/ecop.2004.7349

http://www.persee.fr/doc/ecop_0249-4744_2004_num_163_2_7349

Document généré le 17/06/2016

Abstract

Break Tests : An Application to French Trend GDP by Hervé Le Bihan

This paper presents and applies to GDP recently developed tests for structural change , which allow for the endogenous statistical determination of trend growth break dates and numbers of break-points . These tests serve to investigate the existence of breaks in French trend GDP from 1963 to 2001 using both a segmented trend approach and a difference-stationary approach . Only one significant break (the 1974 slowdown) is found . Consequently , trend growth is evaluated at 5.1 % prior to 1974 and 2.1 % after 1974 .

Résumé

Le présent article décrit des procédures de test de changements structurels développées récemment et propose une application de ces tests au Produit Intérieur Brut . En particulier , dans ces tests , les dates et le nombre de ruptures dans le taux de croissance du PIB tendanciel ne sont pas supposés connus a priori et sont déterminés de façon endogène par une procédure statistique . Les résultats des tests menés sur le PIB français sur la période 1963-2001 conduisent à ne retenir comme significative sur la période qu ' une seule rupture , celle de 1974 , contemporaine du premier choc pétrolier . L ' évaluation de la croissance tendancielle associée à cet exercice est de 5,1 % avant 1974 et de 2,1 % après 1974 .

Tests de rupture : une application au PIB tendanciel français

Hervé Le Bihan^(*)

Le présent article décrit des procédures de test de changements structurels développées récemment et propose une application de ces tests au Produit Intérieur Brut. En particulier, dans ces tests, les dates et le nombre de ruptures dans le taux de croissance du PIB tendanciel ne sont pas supposés connus a priori et sont déterminés de façon endogène par une procédure statistique.

Les résultats des tests menés sur le PIB français sur la période 1963-2001 conduisent à ne retenir comme significative sur la période qu'une seule rupture, celle de 1974, contemporaine du premier choc pétrolier. L'évaluation de la croissance tendancielle associée à cet exercice est de 5,1% avant 1974 et de 2,1 % après 1974.

(*) Centre de Recherche, DGEI/ DEER, Banque de France.
E-mail: herve.lebihan@banque-france.fr

Cet article est une version révisée d'une contribution à la note d'Etudes et de Recherche n° 89 de la Banque de France (Baghli *et alii*, 2002). L'auteur remercie vivement P. Malgrange (rédacteur en chef de la revue), deux rapporteurs anonymes ainsi que S. Avouyi-Dovi, G. Cette, E. Jondeau, J. Matheron et les participants à un séminaire de la Banque de France, pour leurs remarques et suggestions. Il demeure seul responsable des erreurs et insuffisances subsistantes. Ce texte n'exprime pas nécessairement la position de la Banque de France.

Historiquement, une des premières approches utilisées pour représenter et évaluer le produit intérieur brut (PIB) tendanciel est la tendance «segmentée». Cette méthode a ensuite été abondamment critiquée dans les années quatre-vingt pour deux types de raisons, développées par exemple dans Doz, Rabault et Sobczak (1995). D'une part, de nombreux travaux relevant de l'économétrie des variables non-stationnaires ont suggéré que le PIB des pays industriels était mieux représenté par une tendance stochastique que par une tendance déterministe. D'autre part, dans l'approche par la tendance segmentée, le nombre et la datation des ruptures sont traditionnellement fondés sur dires d'experts, et traités comme connus *a priori* dans l'application des tests de rupture. Les résultats obtenus dépendent ainsi de choix préalables et potentiellement arbitraires.

Toutefois, l'économétrie des changements structurels s'est enrichie au cours des dernières années (voir Hansen, 2001, pour un résumé) et propose de nouveaux outils pour aborder la question des ruptures dans le rythme de la croissance tendancielle. En premier lieu, des tests de racine unité contre l'hypothèse alternative de tendance déterministe avec rupture ont été développés, dans la lignée de la contribution initiale de Perron (1989). En second lieu, les tests de rupture les plus récents se dispensent de la connaissance *a priori* des dates de ruptures et de leur nombre. Cette littérature fournit ainsi des méthodes pour donner des réponses de nature statistique aux questions suivantes : y-a-t-il eu des ruptures dans le taux de croissance du PIB tendanciel ? En quel nombre ? À quelles dates se sont-elles produites ?

Dans la présente note, nous présentons successivement une brève revue de ces méthodes (première et deuxième parties) puis leur application au PIB français sur la période 1963-2001 (troisième partie)⁽¹⁾. La datation des ruptures retenue permet de calculer un taux de croissance tendanciel du PIB, évalué sans recours une périodisation *a priori* des cycles.

Par ailleurs, l'exercice mené permet d'apporter un éclairage sur les débats récents relatifs aux fluctuations de la croissance potentielle en France. Par exemple, une hypothèse formulée par différents observateurs est celle d'un ralentissement de la croissance potentielle au début des années quatre vingt-dix, en raison de la baisse du progrès technique ou des effets des mesures en faveur de l'emploi non-qualifié sur la productivité du travail. Une autre hypothèse est celle d'une augmentation de la croissance potentielle dans la période récente, liée à l'apparition de la « nouvelle économie ». L'approche par les ruptures de tendance mise en œuvre dans cet article présente l'avantage, par rapport aux méthodes de type lissage souvent utilisée dans l'analyse de la croissance potentielle, de permettre de tester la significativité des inflexions. Il convient de souligner que, même si tout indicateur de PIB tendanciel peut être envisagé comme une « mesure » de PIB potentiel, la notion de PIB potentiel se distingue de celle de PIB tendanciel et est plus complexe que cette dernière⁽²⁾. De plus, l'approche ici suivie pour caractériser la croissance tendancielle présente plusieurs limites, communes à la plupart des approches statistiques de la croissance tendancielle : elle est strictement univariée ; les résultats obtenus n'ont pas nécessairement une interprétation économique, et peuvent se révéler peu robustes en fin de période.

PIB tendanciel et ruptures de tendance : modélisation

Modèles TS et DS du PIB, et croissance tendancielle

Le présent article s'intéresse à la présence de rupture sur la croissance tendancielle du PIB. Une approche simple pour évaluer le PIB tendanciel consiste à estimer le modèle suivant, selon lequel le logarithme du PIB suit un *trend linéaire* :

$$(1) \quad y_t = \mu + \beta t + v_t$$

où y_t est une observation du logarithme du PIB, t décrit le temps et correspond aux dates d'observation : $t=1 \dots T$. Le paramètre β s'interprète comme le taux de croissance tendanciel. Le terme v_t , résidu de moyenne nulle, est la composante cyclique ou écart de production (*output gap*). Ce terme v_t est vraisemblablement autocorrélé. En revanche v_t est supposé stationnaire : le PIB est alors stationnaire autour d'une tendance. Le modèle (1) en conjonction avec l'hypothèse de stationnarité de v_t est dénommé modèle « TS ».

Cependant la stationnarité autour d'une tendance du PIB est souvent rejetée par l'économétrie des variables non-stationnaires. L'alternative typique est l'hypothèse que le PIB est stationnaire en différence (DS). Cette hypothèse conduit au modèle suivant (marche aléatoire « avec dérive ») :

$$(2) \quad \Delta y_t = b + e_t$$

Dans ce modèle, le terme d'erreur e_t est supposé stationnaire (mais probablement autocorrélé). La constante b est la composante déterministe du taux de croissance du PIB, et peut s'interpréter à ce titre comme le taux de croissance tendanciel. Il convient de souligner que la différence entre modèles TS et DS tient à l'hypothèse faite sur les résidus, plutôt qu'à l'utilisation de l'équation (1) et (2). Une écriture (2) s'obtient en effet simplement par différence première de l'équation (1), et en posant $e_t = v_t - v_{t-1}$ et $b = \beta$. Inversement, le modèle (2) peut être réécrit sous la forme (1) en notant $\mu = y_0$, et $v_t = \sum_{i=0}^{t-1} e_{t-i}$.

L'évaluation de la composante déterministe de la croissance est assez proche dans les deux modèles. En particulier, l'estimation de β ou de b par les moindres-carrés ordinaires (MCO) fournit des résultats convergents et asymptotiquement équivalents, que le processus pour v_t dans l'écriture (1) soit $I(0)$ ou $I(1)$ (voir sur ce point Canjels et Watson, 1997).

En revanche, les modèles TS et DS renvoient à des représentations très différentes du niveau de PIB tendanciel. L'hypothèse de stationnarité autour d'une *trend*, la plus conforme à la notion historique de PIB tendanciel, conduit à définir un écart de production de façon intuitive comme le résidu de l'équation. On peut également calculer un niveau de PIB tendanciel à l'aide de l'équation (2) en utilisant la décomposition de Beveridge et Nelson, ce qui conduit à une évaluation de l'écart de production très différente. Par ailleurs, l'équation (2) indique qu'après un choc, le *niveau* du PIB tendanciel ne revient pas à un niveau prévisible avant le choc.

Ruptures de tendance

Cet article s'intéresse à la présence de ruptures sur la croissance tendancielle, soit le paramètre β ou b , dans les modèles TS et DS. L'approche par les ruptures de tendance consiste à généraliser le modèle (1) ou (2) en autorisant, dans le cas le plus général, m ruptures sur les paramètres, et donc $m+1$ régimes de croissance.

Dans le cas du modèle (1) le modèle général est celui de la tendance segmentée :

$$(3) \quad y_t = \mu_1 + \beta_1 t + v_t \quad \text{pour } t = 1, \dots, T_1$$

$$y_t = \mu_2 + \beta_2 t + v_t, \quad \text{pour } T_1 + 1, \dots, T_2$$

$$y_t = \mu_{m+1} + \beta_{m+1} t + v_t, \quad \text{pour } t = T_m + 1, \dots, T_{m+1} \quad (\text{avec } T = T_{m+1})$$

Tester l'absence de ruptures revient à tester le modèle (1) contre le modèle (3), plus général. Une rupture sur le modèle TS peut prendre plusieurs formes particulières intéressantes. Nous les détaillons ici en utilisant la typologie et la terminologie de Perron (1989) qui distingue trois modèles A, B, C. Pour simplifier nous exposons le cas où le modèle connaît une seule rupture à la date T_1 . Par ailleurs, dans les différents cas A, B, C, le modèle pour le premier régime de croissance est supposé s'écrire :

$$y_t = \mu_1 + \beta_1 t + v_t \quad \text{pour } t = 1, \dots, T_1$$

Le modèle A est celui d'un changement permanent du niveau du PIB tendanciel à la date T_1 (un « crash »). Pour la deuxième période ($t = T_1 + 1, \dots, T$) le modèle est donc :

$$y_t = \mu_2 + \beta_1 t + v_t$$

et $\mu_2 - \mu_1$ est le saut du PIB tendanciel à la date $T_1 + 1$.

Le modèle B traduit un changement de pente du PIB, avec un raccordement continu en $t = T_1$ dans le niveau de la composante déterministe. En utilisant la variable indicatrice $I(\cdot)$, pour T_1 donné, le modèle s'écrit sur toute la période :

$$(4) \quad y_t = \mu_1 + \beta_1 t + (\beta_2 - \beta_1)(t - T_1)I(t > T_1) + v_t, \quad t = 1, \dots, T$$

Notons que, sur la seule seconde période, le modèle s'écrit bien $y_t = \mu_2 + \beta_2 t + v_t$ conformément à l'équation (3), les paramètres étant alors liés par la relation $\mu_2 = \mu_1 + (\beta_2 - \beta_1)T_1$.

Le modèle C est le cas général décrit par (3) pour $m = 1$: un changement de pente et de niveau du PIB. Sur la seconde période, $y_t = \mu_2 + \beta_2 t + v_t$, sans contrainte entre les paramètres. Dans l'application qui suit, nous choisissons, au sein des modèles TS, de privilégier le modèle (B), en particulier en raison de la difficulté

d'interpréter des sauts à la hausse ou à la baisse du PIB tendanciel. Considérer des ruptures de niveau serait toutefois nécessaire lors de chocs majeurs, comme des changements institutionnels (la réunification, par exemple, dans le cas de l'Allemagne), des guerres (dans une analyse de très longue période), etc.

La spécification DS permet de la même façon de modéliser et de tester des ruptures dans la composante déterministe de croissance de long terme, matérialisées par un changement du paramètre b . Banerjee, Lumsdaine et Stock (1992), par exemple, retiennent des ruptures sous cette forme pour différents pays industriels. Dans le cas d'une seule rupture le modèle, sous l'hypothèse alternative d'instabilité, est :

$$(5) \Delta y_t = b_1 + (b_2 - b_1)I(t > T_1) + v_t, \quad t=1, \dots, T$$

On dénommera ce dernier modèle, modèle D.

Autocorrélation des résidus

L'autocorrélation des résidus doit être discutée en raison de son importance pour l'estimation et les tests de rupture. L'autocorrélation des résidus est une caractéristique vraisemblable des modèles TS (équations (1) et (4)), car le résidu s'interprète comme la composante cyclique du PIB. Cette variable présente une importante persistance en raison de la persistance des chocs qui affectent l'économie, et des délais inhérents à la propagation de ces chocs. Les modèles DS (équation (2) et (5)) présentent également une autocorrélation des résidus car le passage à la différence première ne suffit pas à capter toute la dynamique du produit intérieur brut.

La présence de cette autocorrélation dans les résidus ne biaise pas l'estimation des paramètres de ces différents modèles. Mais elle peut affecter la validité des tests de rupture. Deux approches concurrentes sont possibles pour traiter la présence d'autocorrélation des erreurs. La première, que nous dénommons ici approche non-paramétrique, consiste à estimer un modèle statique comme (1) ou (2) par les moindres carrés ordinaires, puis à prendre en compte l'autocorrélation lors de la réalisation de tests de rupture, en utilisant une estimation robuste de la matrice de variance des paramètres (la procédure de Newey-West, 1987, par exemple).

La seconde approche (dénommée ici approche paramétrique) consiste à effectuer des régressions augmentées, *i.e.* à ajouter dans l'équation (1) des termes autorégressifs sous la forme :

$$(6) y_t = \mu_1^* + \beta_1^* t + c_1 y_{t-1} + \dots + c_p y_{t-p} + \varepsilon_t$$

Les résidus peuvent alors être considérés comme des bruits blancs. Cette procédure se justifie particulièrement lorsque l'autocorrélation des résidus prend la forme d'un processus autorégressif (AR). Ainsi considérons le cas où, dans l'équation (1), les résidus suivent un processus AR(1) d'innovation ε_t :

$$v_t = \rho v_{t-1} + \varepsilon_t$$

La quasi-différenciation de l'équation (1) donne :

$$y_t = \mu(1-\rho) - \rho\beta + \beta(1-\rho)t + \rho y_{t-1} + \varepsilon_t$$

En posant $c = \rho$, $\mu^* = \mu(1-\rho) - \rho\beta$, $\beta^* = \beta(1-\rho)$, on retrouve bien une écriture (6) dont le résidu est un bruit blanc. L'estimation d'une équation comme (6) s'apparente dans ce cas à l'application de moindres-carrés généralisés.

Dans la pratique, l'approche paramétrique requiert une procédure de sélection du nombre de retards p à introduire dans l'équation. Il convient de relever que tester la stabilité du modèle (6) peut prendre différents sens. En effet on peut tester une rupture partielle limitée aux paramètres de pente et de niveau (μ^* et β^*), ou une rupture de l'ensemble des paramètres y compris ceux décrivant la dynamique (les paramètres d'autocorrélation c_i). Dans l'exercice qui suit, nous effectuons des tests de rupture partielle (portant sur le paramètre de *trend*), suivant en cela plusieurs études antérieures (Perron, 1989, Lumsdaine et Papell, 1997, Vogelsang, 1997). Si cette restriction constitue une limite à l'exercice proposé, notons que l'approche non paramétrique est elle robuste à l'autocorrélation et l'hétéroscédasticité qu'introduirait une rupture sur les paramètres autorégressifs.

Sous l'hypothèse de rupture le modèle estimé est donc dans le cas TS :

$$(7) y_t = \mu_1^* + \beta_1^* t + \gamma(t - T_1)I(t > T_1) + \sum_{i=1}^p c_i y_{t-i} + \varepsilon_t$$

Dans le cas de la modélisation DS, l'approche paramétrique se fonde de la même façon sur une équation autorégressive :

$$(8) \Delta y_t = b_1^* + (b_2^* + b_1^*) I(t > T_1) + \sum_{i=1}^p c_i \Delta y_{t-i} + u_t$$

La croissance tendancielle de long terme se déduit de l'équation estimée respectivement par $\beta_1 = \beta_1^* / (1 - \sum_{i=1}^p c_i)$ dans le cas TS et $b_1 = b_1^* / (1 - \sum_{i=1}^p c_i)$ dans le cas DS. Il convient de relever que l'on ne peut aisément hiérarchiser dans l'absolu l'approche par MCO statique et l'approche autorégressive. Leur performance relative en petit échantillon pour l'estimation des paramètres de croissance tendancielle dépend de la forme de l'autocorrélation effectivement présente dans le modèle⁽³⁾. Nous renvoyons sur ce point à l'article de Canjels et Watson (1997).

Notons enfin que selon l'approche choisie (paramétrique ou non-paramétrique), le modèle estimé - (4) ou (7) dans le cas TS, (5) ou (8) dans le cas DS- présente des propriétés dynamiques différentes, consécutivement à une rupture sur les paramètres. Lorsque la rupture affecte le PIB, la transition vers le nouveau PIB tendanciel est instantanée dans le cas de la correction non-paramétrique, car la dynamique du modèle ne concerne que les résidus (le modèle est de type *additive outlier*). Elle est au contraire graduelle dans le cas de l'approche paramétrique (le modèle est de type *innovational outlier*).

Non-stationnarité stochastique versus tendance segmentée : discussion

Pour tester et identifier les ruptures sur le rythme de croissance tendancielle, une première stratégie consiste à sélectionner une spécification, (1) ou (2), à l'aide de tests préliminaires sur la forme de la non-stationnarité. Le choix entre l'hypothèse de stationnarité en différence et celle de stationnarité autour d'un trend fait l'objet de controverses, empiriques et théoriques, qui durent depuis de nombreuses années. Une contribution importante à ce débat est celle de Perron (1989) qui propose des tests de l'hypothèse de racine unité – de type ADF (Dickey-Fuller augmentés) notamment - autorisant sous l'hypothèse alternative une tendance déterministe avec rupture⁽⁴⁾. Depuis, Zivot et Andrews (1992), Banerjee, Lumsdaine et Stock (1992) et Perron (1997) ont étendu ces tests en permettant l'endogénéisation de la date de rupture. Ces tests sont les plus pertinents dans notre contexte.

Nous décrivons ici la version de ces tests proposée par Perron (1997), utilisée dans l'application *infra*. Elle consiste à estimer pour toutes les dates de ruptures possibles, une équation identique à (4) ci-dessus, et à effectuer dans chaque cas une régression de type ADF sur les résidus estimés \tilde{v}_t , soit :

$$\tilde{v}_t = \alpha \tilde{v}_{t-1} + \sum_{i=1}^K c_i \Delta \tilde{v}_{t-i} + e_t$$

À chaque date de rupture, le choix du nombre de retards K est effectué selon une démarche descendante. L'équation est d'abord estimée avec un nombre de retard maximal ($K = K_{\max}$). Si le dernier retard n'est pas significatif au seuil de 10% (lorsque le t de Student est inférieur à 1.6), l'équation est ré-estimée avec ($K_{\max} - 1$) retards, et on élimine séquentiellement le dernier retard jusqu'à obtention d'un modèle où le dernier retard est significatif. La statistique de test est ensuite la statistique de Student associée à l'hypothèse $\alpha = 1$, pour la date de rupture optimale. En pratique, cette date de rupture est déterminée soit comme celle qui minimise la statistique ADF elle-même, soit comme celle qui maximise la valeur absolue de la statistique de Student sur le paramètre $\gamma = \beta_2 - \beta_1$. La statistique ADF calculée à l'aide du premier critère sera notée t_α et celle calculée à l'aide du second critère $t_{\alpha, \gamma}$. L'hypothèse de racine unité est rejetée si la statistique dépasse en valeur absolue des valeurs critiques tabulées par Perron (1997).

L'utilisation de tests de racine unité soulève différentes difficultés. D'une part, ces tests ont notoirement une faible puissance en échantillon fini. En outre, les résultats des tests de racine-unité se révèlent parfois sensibles à des choix de spécification relativement arbitraires. C'est le cas dans l'application menée ci-dessous. Ces arguments ne discréditent néanmoins pas les résultats de tests de racine unité, car de nombreux tests, comme les tests de stabilité présentés dans la section qui suit, peuvent se heurter à des difficultés similaires en petits échantillons.

Une difficulté supplémentaire est qu'en toute rigueur on devrait tester simultanément la présence de rupture et la stationnarité du modèle. En effet, tester la stationnarité préalablement à l'identification des ruptures est

problématique car la présence de ruptures affecte les tests de stationnarité. Toutefois, à notre connaissance, il n'existe pas à ce jour de procédure permettant de mener simultanément test de rupture et test de stationnarité.

Pour contourner ces difficultés, nous retenons dans l'application l'approche pragmatique suivante : tester la présence de rupture à la fois sous l'hypothèse de *trend*-stationnarité (« TS », modèle B) et sous l'hypothèse de stationnarité en différence (modèle D). La confrontation des résultats tirés de ces deux modèles fournit une indication sur la robustesse de la datation et de la significativité des ruptures de tendance. En particulier, si les conclusions des tests de stabilité se révèlent analogues lorsqu'il sont menés en utilisant deux spécifications concurrentes, les résultats sur la question de la stabilité peuvent être regardés comme fiables⁽⁵⁾.

On doit bien sûr garder à l'esprit que, pour chaque test réalisé, au moins l'une des deux équations TS ou DS est mal spécifiée, car les deux représentations de la non-stationnarité ne sont pas compatibles. Cet argument peut conduire à critiquer la démarche adoptée ici, et la meilleure démarche à adopter en cas d'incertitude sur le degré d'intégration du PIB demeure sujette à controverse. Cependant, un argument en faveur de l'approche ici retenue est à nos yeux que l'identification des ruptures sur la composante déterministe est robuste à l'hypothèse faite sur le degré d'intégration de la série. L'annexe 1 en témoigne, en présentant des simulations de Monte Carlo illustratives : la date de rupture est en effet estimée sans biais, même lorsque la spécification utilisée pour le test correspond pas au vrai modèle. Des propriétés analytiques établies par Perron et Zhu (2002) confortent ces résultats.

Tester des ruptures à des dates inconnues

Cette section passe en revue les procédures permettant de tester la stabilité de la composante déterministe de la croissance du PIB dans les modèles décrits précédemment. Partant de la statistique de Chow traditionnelle, nous décrivons des procédures plus récentes qui visent à tester l'hypothèse de stabilité contre l'hypothèse alternative de ruptures multiples, situées à des dates inconnues. Notons que les tests décrits dans cette section ne sont pas spécifiques à la problématique du PIB tendanciel, ni à des modélisations de type (1) ou (2) : ils peuvent être appliqués à tout type de régression linéaire sur séries temporelles, pour tester la stabilité des paramètres.

Test d'une rupture

Tester la pertinence d'une rupture revient à tester la stabilité de l'équation (1) ou (2). Lorsque l'hypothèse alternative est celle d'une seule rupture ($m = 1$) à une date donnée et connue T_1 , le traditionnel test de Chow (1960) peut être appliqué. La statistique de test s'écrit :

$$F_T^{Chow}(T_1) = \frac{(T-2q-p)}{q} \left(\frac{SSR_{1,T} - SSRA_{T_1}}{SSRA_{T_1}} \right)$$

$SSR_{1,T}$, $SSRA_{T_1}$ représentent respectivement la somme des carrés des résidus pour le modèle sans rupture estimé pour les dates 1 à T , et pour le modèle autorisant une rupture sur un ou plusieurs paramètres à la date T_1 ; q est le nombre de paramètres sujets à rupture ; p est le nombre de paramètres supposés stables. La statistique suit une loi de Fisher à q et $T - 2q - p$ degrés de liberté à distance finie si les erreurs sont indépendantes identiquement distribuées (*i.i.d*) et suivent une loi normale. Dans la suite nous utilisons la forme de Wald du test de Chow, valable sous des hypothèses moins restrictives. Elle est fondée sur la statistique $F = q F^{Chow}$ et est donc donnée par :

$$(9) \quad F_T\left(\frac{T_1}{T}\right) = (T-2q-p) \left(\frac{SSR_{1,T} - SSRA_{T_1}}{SSRA_{T_1}} \right)$$

Cette statistique suit asymptotiquement une loi du Chi-2 à q degrés de liberté, sous l'hypothèse d'erreurs indépendantes identiquement distribuées, sans restriction sur la loi suivie. Une écriture de la statistique de test numériquement équivalente à (9) est la suivante :

$$(10) \quad F\left(\frac{T_1}{T}\right) = \frac{(T-2q-p)}{T} \hat{\delta}' R' [R' \hat{V}(\hat{\delta}') R']^{-1} R \hat{\delta}$$

où $\hat{\delta}$ est le vecteur des paramètres estimés, R une matrice telle que sous l'hypothèse nulle de stabilité $R\delta=0$, et $\hat{V}(\hat{\delta})$ est la variance estimée de $\hat{\delta}$. Dans le cas d'erreurs *i.i.d.* $\hat{V}(\hat{\delta})=\hat{\sigma}^2 (X'X)^{-1}$ où $\hat{\sigma}^2$ est un estimateur de la variance des résidus u_t et X est la matrice des régresseurs empilés. Par exemple dans le cas du modèle B (équation (4)), la matrice X est constituée des vecteurs lignes $x'_t=(1, t, (t-T_1).I(t>T_1))$, $\delta=(\mu_1, \beta_1, \gamma)'$ et $R=(0,0,-1)$. Un intérêt de l'expression (10) est qu'elle permet de traiter le cas d'erreurs autocorrélées : $\hat{V}(\hat{\delta})$ est alors obtenue en appliquant une procédure de calcul de la matrice de variance-covariance robuste à l'auto-corrélation. Cette variance est donnée par $\hat{V}(\hat{\delta})=(X'X)^{-1} \hat{S} (X'X)$ où \hat{S} est un estimateur de la variance de long terme du vecteur $x_t u_t$. Dans le cas de la procédure de Newey-West (1987) :

$$(11) \hat{S}=\hat{\Gamma}_0+\sum_{i=1}^M\left[1-\frac{i}{M+1}\right](\hat{\Gamma}_i+\hat{\Gamma}_i')$$

$$\text{et } \hat{\Gamma}_i=(1/T)\sum_{t=M+1}^T x_t \hat{u}_t \hat{u}_{t-i}' x_{t-i}'$$

Le paramètre M est la taille de la « fenêtre » utilisée pour calculer les autocovariances.

Lorsque la date de rupture T_1 n'est pas connue, une idée naturelle est de calculer la statistique F_T pour l'ensemble des valeurs possibles de T_1 , et de retenir celle qui fournit la plus grande valeur de F_T . Ceci équivaut à sélectionner la date de rupture qui minimise la somme des carrés des résidus⁽⁶⁾. La statistique ainsi définie est appelée « Sup F » :

$$(12) \text{SupF} = \underset{T_1 = T_{\min}, \dots, T_{\max}}{\text{Max}} F\left(\frac{T_1}{T}\right)$$

Le calcul de la statistique SupF fournit ainsi un estimateur de la date de rupture T_1 la plus plausible⁽⁷⁾. Toutefois pour juger si la rupture identifiée est significative, la distribution usuelle du test de Chow (*i.e.* la loi du Chi-deux sous la forme de Wald du test) n'est plus valide et doit être remplacée par une distribution non-standard tabulée par Andrews (1993). Les seuils critiques pour accepter l'hypothèse de rupture sont *plus élevés* que ceux fournis par la table statistique du Chi-deux usuelle (on rejettera moins facilement l'hypothèse de stabilité qu'avec les tables usuelles). En effet, par construction, la statistique SupF est supérieure (ou égale) à celle obtenue avec une date de rupture fixe quelconque. La sélection par balayage dans la date de rupture possible constitue un « pré-test ». Utiliser la distribution usuelle, en ignorant le fait que la statistique est obtenue à l'issue de ce « pré-test » pourra conduire parfois à rejeter à tort l'hypothèse de stabilité. En d'autres termes les résultats obtenus en utilisant les valeurs critiques traditionnelles seront biaisés en faveur de l'hypothèse de rupture.

Andrews (1993) a établi la distribution des statistiques SupF pour des régresseurs stationnaires et sans tendance déterministe. Le cas de régresseurs incluant des tendances déterministes a été considéré dans le cas *i.i.d* par Vogelsang (1997). Selon Bai et Perron (1998b), on peut de façon générale utiliser les mêmes valeurs critiques avec et sans tendance déterministe parmi les régresseurs.

Il existe de nombreux autres tests de stabilité ne spécifiant pas de date *a priori* (voir, par exemple, le *survey* de Stock, 1994). Andrews et Ploberger (1994) ont notamment proposé et recommandé, en raison de leurs propriétés d'optimalité, les statistiques ExpF et AvgF qui s'écrivent respectivement :

$$\text{ExpF} = \ln \left[\frac{1}{T_{\max} - T_{\min} + 1} \sum_{T_{\min}}^{T_{\max}} \exp \left(\frac{1}{2} F \left(\frac{T_1}{T} \right) \right) \right]$$

$$\text{AvgF} = \left(\frac{1}{T_{\max} - T_{\min} + 1} \right) \sum_{T_{\min}}^{T_{\max}} F \left(\frac{T_1}{T} \right)$$

Ces deux statistiques suivent asymptotiquement des lois tabulées par Andrews et Ploberger (1994). Elles conduisent à des tests plus puissants que celui fondé sur la statistique SupF. La statistique SupF conserve un intérêt particulier parce qu'elle est fondamentalement liée à l'estimation des dates de rupture, ce qui est l'un des objectifs de notre application, et parce qu'elle a été étendue aux cas de ruptures multiples.

Ruptures multiples

Les tests F et SupF décrits ci-dessus se généralisent au cas de ruptures multiples, si on suppose le nombre de ruptures connu (Bai et Perron, 1998a et b)⁽⁸⁾. Lorsque le nombre de ruptures n'est pas supposé connu, ce nombre doit également être estimé, ce qui soulève des problèmes additionnels.

Ruptures multiples en nombre connu

Lorsqu'à la fois le nombre et les dates de ruptures potentielles sont connus, on peut utiliser le test de Chow traditionnel en amendant de manière adéquate les formules (9) et (10). La statistique de Wald pour m ruptures aux dates T_1, \dots, T_m s'écrit :

$$(13) F(m) = \frac{(T - (m+1)q - p)}{Tm} \hat{\delta}' R' [R' \hat{V}(\hat{\delta}') R]^{-1} R \hat{\delta}$$

Le cas de ruptures multiples à des dates inconnues a été traité récemment. Bai et Perron (1998a et b) et Bai (1999) ont proposé une généralisation des tests SupF. Ces auteurs proposent plusieurs statistiques de tests. La statistique $SupF(m)$ consiste à sélectionner la valeur maximale de toutes les statistiques de test de l'hypothèse de stabilité ($m = 0$) contre l'hypothèse de m ruptures. Un autre test, fondé sur la statistique $supF(m+1|m)$, consiste à fixer m dates de ruptures dans une étape préalable et à retenir comme hypothèse nulle la présence d'exactly m ruptures. La statistique $supF(m+1|m)$ est alors la valeur maximale des statistiques de Wald testant le modèle à m ruptures contre chacun des modèles autorisant $m+1$ ruptures, conditionnellement aux m premières dates de ruptures. Cette valeur est obtenue en balayant sur l'ensemble des dates de ruptures supplémentaires possibles. Ces différentes statistiques suivent des lois non-usuelles tabulées dans Bai et Perron (1998a et b).

Nombre de ruptures inconnu : tests SupF séquentiels

Le cas le plus général, au sens où un minimum d'information *a priori* est supposé connu de l'économetre, est celui où le nombre même de ruptures est inconnu. Bai et Perron (1998a et b) ont proposé une approche séquentielle pour estimer ce nombre. On teste tout d'abord l'hypothèse de stabilité du modèle contre la présence d'une rupture. Si la stabilité est rejetée on impose une rupture, et on teste l'hypothèse nulle d'une rupture contre l'hypothèse alternative de deux ruptures à l'aide de la statistique $supF(m+1|m)$. La procédure est répétée pour m croissant et est arrêtée dès que l'hypothèse de stabilité, conditionnellement à m ruptures ne peut être rejetée.

Deux versions de cette procédure de test séquentiel du nombre de rupture existent en fait. La première (Bai et Perron, 1998a et 1998b) repose sur la détermination séquentielle des points de rupture et la seconde (Bai, 1999) est fondée sur la détermination simultanée des points de rupture. Dans la première version, la date de rupture obtenue à la première étape de la procédure est conservée lorsque l'on teste une seconde rupture. Cette démarche se fonde sur une propriété théorique établie par Bai (1996) : l'estimation séquentielle des ruptures permet une estimation convergente des dates de ruptures.

La seconde approche repose sur l'estimation simultanée de toutes les dates de rupture. Pour chaque nombre m de rupture fixé, les m dates de ruptures (T_1, \dots, T_m) sont sélectionnées librement. Bai (1999) propose un test fondé sur le rapport des vraisemblances des modèles à m et $m+1$ ruptures. Toutefois la distribution de la statistique n'est établie que pour des résidus indépendants identiquement distribués. Nous nous limitons donc ici à la détermination séquentielle des dates de rupture.

Nombre de ruptures inconnu : l'approche par les critères d'information

Une approche plus simple de la détermination du nombre de rupture repose sur le critère d'information BIC (*Bayesian Information Criterion*). L'utilisation de ce critère a été proposée par Yao (1987) et Kim (1997). La méthode consiste à estimer le modèle pour tous les nombres de rupture et toutes les dates de rupture possibles et à sélectionner le modèle dont le critère BIC est minimal. Pour un modèle donné,

$$(14) BIC = \ln\left(\frac{SSR}{T}\right) + [(m+1)q + m + p] \frac{\ln(T)}{T}$$

où m est le nombre de ruptures, q celui de paramètres affectés par une rupture et enfin p le nombre de paramètres non affectés par une rupture⁽⁹⁾.

Le critère BIC correspond ainsi à une fonction de la somme des carrés des résidus du modèle augmentée d'un terme pénalisant le nombre de ruptures. Ce terme contrebalance le fait qu'autoriser des ruptures supplémentaires dans un modèle réduit mécaniquement la somme des carrés des résidus.

Yao (1987) a fourni des fondements statistiques à ce critère en montrant qu'il fournit un estimateur convergent du nombre de ruptures. Toutefois, ce résultat n'est établi que sous des hypothèses très restrictives (normalité et non-autocorrélation des résidus) qui sont très contraignantes pour les applications envisagées ici. Par ailleurs, les simulations menées par Bai et Perron indiquent que la procédure de sélection du nombre de ruptures par l'approche SupF séquentielle est systématiquement plus performante que l'approche par le critère BIC. Dans le cadre du présent exercice, des résultats préliminaires (non rapportés) indiquent que le critère BIC tend à fournir un nombre de ruptures très élevé dans le cas de modèles sans termes retardés comme (4) et au contraire nul dans le cas de modèles autorégressifs comme (7). Nous choisissons donc dans ce travail de présenter exclusivement les résultats obtenus avec les tests SupF, ExpF, AvgF, sans détailler l'approche par les critères d'information.

Intervalles de confiance

Lorsque la stabilité du modèle est rejetée et qu'un nombre de rupture est déterminé par les procédures décrites ci-dessus, la date (ou les dates) de rupture peut être considérée comme un paramètre à estimer. Bai (1997) a dans cette logique proposé un calcul d'intervalle de confiance autour de cette date estimée. Bai (1997) considère le modèle à une rupture :

$$y_t = x_t' \beta + \varepsilon_t, \quad t=1, \dots, T_1$$

$$y_t = x_t' \beta + z_t' \delta + \varepsilon_t, \quad t=T_1+1, \dots, T$$

où les régresseurs z_t sont le sous-ensemble des régresseurs x_t dont le paramètre est affecté par une rupture (il y a rupture de tous les paramètres si $z_t = x_t$). Dans le cas général d'autocorrélation et hétéroscédasticité des résidus, et pour une rupture unique, la procédure est la suivante. On note $\hat{Q}_1 = \frac{1}{T_1} \sum_{t=1}^{T_1} z_t z_t'$ et $\hat{\Omega}_1$ un estimateur de la variance de long terme du vecteur $\varepsilon_t z_t$ pour la première période (*i.e.* pour les observations $t=1$ à T_1), construit de la même façon que \hat{S} dans (11) ci-dessus.

\hat{Q}_2 et $\hat{\Omega}_2$ sont les matrices correspondantes définies pour les observations $t=T_1+1$ à T . On définit également les scalaires

$$\hat{L} = \frac{(\hat{\delta}' \hat{Q}_1 \hat{\delta})^2}{(\hat{\delta}' \hat{\Omega}_1 \hat{\delta})}, \quad \xi = \frac{\hat{\delta}' \hat{Q}_2 \hat{\delta}}{\hat{\delta}' \hat{Q}_1 \hat{\delta}} \quad \text{et} \quad \phi = \frac{\hat{\delta}' \hat{\Omega}_2 \hat{\delta}}{\hat{\delta}' \hat{\Omega}_1 \hat{\delta}}$$

Un intervalle de confiance à 90% autour de la date de rupture est alors donné par :

$$[T_1 - [c_2 / \hat{L}] - 1, T_1 + [c_1 / \hat{L}] + 1]$$

où c_1 et c_2 sont les quantiles à 5% et 95% d'une distribution paramétrée par ξ et ϕ et décrite dans Bai (1997). La procédure ci-dessus ne contraint pas l'intervalle de confiance à être symétrique. Elle suppose que tous les régresseurs sont stationnaires ce qui exclut le cas de tendance linéaire.

Dans le cas du modèle avec *trend* segmenté (équation (4)) Perron et Zhu (2002) montrent que la date de rupture estimée suit une distribution normale :

$$\frac{\hat{T}_1 - T_1^0}{\sqrt{T}} \rightarrow N\left(0, \frac{2\sigma^2}{15(\gamma)^2}\right)$$

Où T_1^0 est la date de rupture, $\gamma = \beta_2 - \beta_1$, σ^2 est la variance de long terme des résidus qui peut être estimée en utilisant la procédure de Newey-West. Cette distribution peut être utilisée pour construire un intervalle de confiance autour de la date de rupture dans ce modèle.

Application au PIB français

Données et spécifications

Les estimations sont menées en utilisant le PIB trimestriel français sur la période 1963:1-2001:4. La série de PIB utilisée est celle des comptes trimestriels, base 1995, dans leur livraison de mai 2002. Une série de longue période de PIB est constituée par rétropolation à l'aide du PIB base 1980 (sur la période 1970:1-1977:4) et du PIB base 1970 (sur la période 1963:1 à 1969:4). La rétropolation est effectuée en utilisant le taux de croissance des bases 1970 et 1980. Afin d'estimer le modèle sur la même période 1963:1-2001:4, lorsqu'un modèle auto-régressif est estimé, nous rétopolons également le PIB en utilisant une série historique de PIB marchand, pour disposer des p trimestres antérieurs à 1963:1⁽¹⁰⁾. Le nombre d'observations est égal à $T=156$ pour toutes les régressions. Enfin, le second trimestre 1968 est corrigé de l'effet des grèves en interpolant les données du premier et du troisième trimestre⁽¹¹⁾.

Le tableau 1 présente les résultats de tests de racine unité (contre une tendance segmentée présentant une rupture à date inconnue) à la Perron (1997) menés sur cette série. Nous retrouvons ici le conflit obtenu dans le cas de différents pays industriels dont la France entre Banerjee, Lumsdaine et Stock (1992) et Perron (1997). En effet si l'on retient $K_{\max}=8$ l'hypothèse de racine unité ne peut être rejetée. En revanche avec $K_{\max}=13$ ⁽¹²⁾ cette hypothèse est rejetée en faveur du modèle TS avec rupture au seuil de 10% et 5% respectivement pour les statistiques t_{α} et $t_{\alpha,\gamma}$. Notre lecture de ce résultat est qu'il demeure difficile de trancher entre l'hypothèse de *trend*-stationnarité avec rupture et l'hypothèse de stationnarité en différence, même si cette dernière est le plus souvent adoptée dans les travaux empiriques récents. Il nous semble dès lors indiqué de considérer les tests de rupture sous chacune des hypothèses alternativement.

Quatre spécifications sont retenues pour mener des tests de rupture : les modèles TS et DS, estimés avec et sans correction paramétrique. Pour m ruptures situées aux dates T_1, \dots, T_m , le modèle TS s'écrit sans correction paramétrique :

$$(15) \quad y_t = \mu_1 + \beta_1 t + \sum_{k=1}^m \beta_{k+1} (t - T_k) I(t > T_k) + v_t$$

Et dans le cas d'une correction paramétrique :

$$(16) \quad y_t = \mu_1^* + \beta_1^* t + \sum_{k=1}^m \beta_{k+1}^* (t - T_k) I(t > T_k) + \sum_{i=1}^p c_i y_{t-i} + v_t$$

Pour le modèle DS (stationnarité en différence avec rupture dans la dérive) les deux spécifications sont :

$$(17) \quad \Delta y_t = \beta_1 + \sum_{k=1}^m \beta_{k+1} I(t > T_k) + v_t$$

et dans le cas d'une correction paramétrique :

$$(18) \quad \Delta y_t = \beta_1^* + \sum_{k=1}^m \beta_{k+1}^* I(t > T_k) + \sum_{i=1}^p c_i \Delta y_{t-i} + v_t$$

Tableau 1 : tests de racine-unité contre une tendance avec rupture à date inconnue

	Statistique ADF-Min	K^*
Statistique t_{α}		
$K_{\max} = 8$	-3,44	5
$K_{\max} = 13$	-4,60 *	11
Statistique $t_{\alpha,\gamma}$		
$K_{\max} = 8$	-3,38	5
$K_{\max} = 13$	-4,49 **	11

Note : période d'estimation 1963:1-2001:4 ; */**/** : respectivement significatif à 10%, 5%, 1% ; les valeurs critiques utilisées sont respectivement -4,48/-4,85/-5,45 pour la statistique t_{α} et -3,99/-4,36/-5,00 pour la statistique $t_{\alpha,\gamma}$.
Source : Perron (1997), tableau I. K^* est le retard le plus élevé du modèle sélectionné par la procédure "Kmax".

Tableau 2 : tests de stabilité SupF, ExpF, AvgF

Approche non paramétrique	Modèle TS	Modèle DS
SupF	Eq. (15) 964,82 ***	Eq. (17) 63,85 ***
Date	1973:4	1974:1
ExpF	92,06 ***	8,69 ***
AvgF	163,39 ***	17,31 ***
Approche paramétrique	Modèle TS	Modèle DS
SupF	Eq. (16) 8,75 *	Eq. (18) 22,51 ***
Date	1972:4	1974:1
ExpF	2,05 *	7,98 ***
AvgF	1,79	6,70 ***

Note : Période : 1963:1-2001:4. */**/* : respectivement significatif à 10%, 5%, 1% ; les valeurs critiques utilisées sont pour SupF : 7,63/9,31/12,69 ; pour ExpF : 1,52/2,08/3,41 ; pour AvgF : 2,09/2,77/4,43 (sources : Andrews, 1993, tableau 1 avec $m=1$ paramètre sujet à rupture – soit $p=1$ dans la notation d'Andrews-, et Andrews et Ploberger, 1994). Date : date de rupture estimée par la procédure SupF.

Les tests effectués sont les tests SupF, ExpF, AvgF, où F représente la statistique de Wald testant la nullité des paramètres $\beta_k, k>1$, sous l'hypothèse de stabilité des paramètres μ et c_i .

Dans l'approche paramétrique, le choix du paramètre p est effectué selon la démarche descendante préconisée par Perron (1989) et décrite *supra*, appliquée au modèle sans rupture. En débutant avec un nombre de retard maximal $p_{\max}=8$, nous obtenons sur la période un paramètre p égal à 5 pour les modèles *trend*-stationnaires et à 4 pour les modèles en différence⁽¹³⁾.

Dans l'approche non-paramétrique, nous utilisons la procédure de Newey-West (1987) pour calculer la matrice de variance de long terme des paramètres $\hat{V}(\hat{\delta})$. Cette matrice est calculée en utilisant une fenêtre de type Bartlett sur $M=4$ trimestres. Précisons enfin qu'un intervalle minimal entre deux ruptures doit être imposé. Nous le fixons ici à 12 trimestres soit environ 8 % de la taille de l'échantillon. L'hypothèse d'une rupture au cours des 12 premières observations ou des 12 dernières est en particulier exclue *a priori*. Les valeurs critiques dépendent de la fraction ϵ des observations ainsi exclues des dates de rupture potentielles.

Résultats

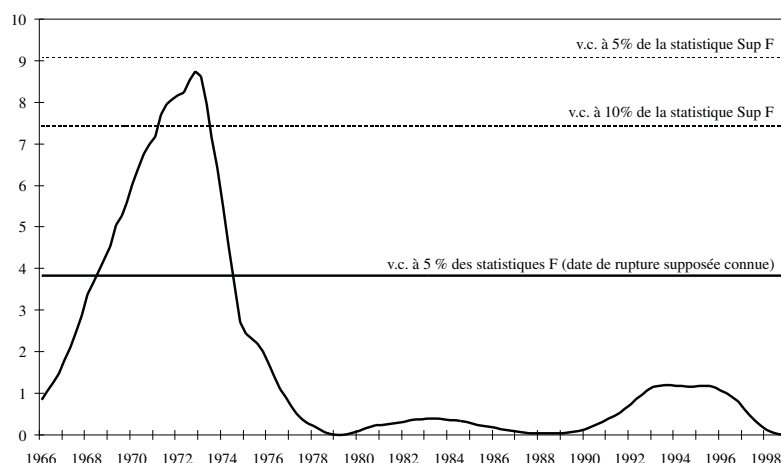
Les résultats des tests ExpF, AvgF et des différents tests de type SupF, menés à partir des estimations sur la période 1963:1-2001:4, sont présentés dans les tableaux 2 et 3. Nous utilisons les seuils critiques tabulés par Andrews (1993), Andrews et Ploberger (1994) et Bai et Perron (1998b) pour une fraction $\epsilon=(T_{\min}/T)=10\%$ de dates exclue *a priori*⁽¹⁴⁾.

Le test d'Andrews SupF (tableau 2) mené sur les modèles (15) à (18), rejette dans chaque cas la stabilité du modèle. Le rejet est très net dans le cas des modèles avec correction non paramétrique de l'autocorrélation. Pour le modèle autorégressif *trend*-stationnaire (équation 16), la statistique de test vaut 8,75 et l'hypothèse de stabilité n'est rejetée qu'au seuil de 10 %. Ceci est illustré par le graphique 1, qui présente la séquence des statistiques F en fonction de la date de rupture supposée et permet d'identifier visuellement la statistique SupF. La date de rupture obtenue est située fin 1972. À titre illustratif, le graphique représente également la valeur critique associée au seuil du test de Chow traditionnel, supposant connue la date de rupture. Pour la date de rupture « optimale », on observe ainsi un conflit d'inférence, au seuil de 5%, selon que la date de rupture est supposée connue (ce qui conduit à rejeter la stabilité) ou que le caractère endogène de la sélection de la date de rupture est reconnu (ce qui conduit au seuil de 5% à ne pas rejeter la stabilité).

Pour toutes les spécifications, la date de rupture est voisine du premier choc pétrolier. Par exemple, dans le cadre du modèle estimé sans termes autorégressifs, la date de rupture estimée est 1973 :4, ce qui indique, dans la notation retenue ici, une rupture entre le quatrième trimestre 1973 et le premier trimestre 1974. Les résultats des tests ExpF et AvgF confirment ceux obtenus par le test SupF. En effet la stabilité est rejetée au seuil de 1% par ces deux tests dans tous les cas, sauf dans celui du modèle *trend*-stationnaire avec correction paramétrique (équation (16)) : la stabilité est alors rejetée seulement au seuil de 10% par la statistique ExpF et n'est pas rejetée par la statistique AvgF.

Un résultat notable du tableau 2 est que les tests menés avec correction paramétrique rejettent moins nettement la stabilité que les tests utilisant la procédure non-paramétrique de correction de l'autocorrélation (*Newey-West*). Ce résultat s'observe également pour les tests présentés dans les tableaux 3 et 4 et discutés ci-dessous, et semble

Graphique 1 : séquence des statistiques de Chow et statistique SupF (modèle autorégressif)



refléter une propriété assez générale. Dans l’annexe 2, nous réalisons différentes simulations de Monte Carlo qui montrent en effet que, dans un modèle avec rupture sur la composante déterministe et autocorrélation des erreurs, la procédure paramétrique détecte moins souvent des ruptures que la procédure non paramétrique (de la même façon elle indique moins souvent à tort une rupture lorsque le vrai modèle est stable). Notre interprétation est la suivante : la correction paramétrique, au travers des termes autorégressifs qu’elle autorise, introduit des degrés de liberté supplémentaires qui permettent d’ajuster la dynamique de la série au moment de la rupture, et rendent une rupture de constante plus difficile à détecter. Un diagnostic plus approfondi sur cette question dépasse toutefois le cadre de cet article.

Le tableau 3 présente les différentes statistiques de type SupF, pour un nombre de ruptures $m=1$ et 2. La statistique SupF(1) coïncide avec la statistique SupF du tableau 1. La statistique SupF(2) teste la stabilité contre la présence de 2 ruptures. Comme la statistique SupF, elle rejette l’hypothèse de stabilité, au seuil de 1 % sauf dans le cas du modèle *trend*-stationnaire autorégressif pour lequel la valeur de la statistique (4,82) ne permet pas de rejeter l’hypothèse de stabilité. Globalement, ces résultats renforcent la présomption de rupture. Ils ne permettent néanmoins pas de conclure à l’existence effective de 2 ruptures. En effet, si la première rupture

Tableau 3 : tests de stabilité contre des ruptures multiples

	Modèle TS	Modèle DS
Approche non paramétrique	Eq. (15)	Eq. (17)
Test de $m=0$ vs. $m=1$ rupture		
Statistique SupF	964,82***	63,85***
Dates	1973:4	1974:1
Significativité		
Test de $m=0$ vs. $m=2$ ruptures		
Statistique SupF(2)	892,51***	33,26***
Dates	1973:1, 1979:2	1974:1, 1997:1
Significativité		
Test de $m=1$ vs. $m=2$ ruptures		
Rupture sous H_0	1973:4	1974:1
Statistique SupF(2 1)	2,2	2,42
Date de la rupture additionnelle	1990:2	1997:2
Approche paramétrique	Eq. (16)	Eq.(18)
Test de $m=0$ vs. $m=1$ rupture		
Statistique SupF(1)	8,75*	22,51***
Dates	1972:4	1974:1
Test de $m=0$ vs. $m=2$ ruptures		
Statistique SupF(2)	4,82	12,40***
Dates	1972:4, 1989:1	1974:1, 1996:4
Test de $m=1$ vs. $m=2$ ruptures		
Rupture sous H_0	1972:04:00	1974:01:00
Statistique SupF(2 1)	0,9	2,14
Date de la rupture additionnelle	1989:1	1996:4

Note : Période : 1963:1-2001:4. */**/** : respectivement significatif à 10%, 5%, 1% ; les valeurs critiques utilisées sont pour SupF : voir tableau 2, pour SupF(2) : 6,93/7,92/10,14, pour SupF(2|1) : 9,97/11,36/15,44 (sources : Bai et Perron 1998 tableau 1). Dates : date(s) de rupture présumée(s).

identifiée se situe entre 1972:4 et 1974:1 pour tous les tests, la datation de la seconde rupture est peu robuste au test effectué. Ainsi, la seconde rupture est localisée au deuxième trimestre 1979 selon le test effectué sur le modèle TS dans l'approche paramétrique, au premier trimestre 1989 selon le test effectué sur le modèle TS dans l'approche non-paramétrique, et en 1996-1997 selon les deux tests effectués sur le modèle DS (modèle stationnaire en différence).

Ce diagnostic d'absence d'une seconde rupture est confirmé par les tests SupF(2|1). En effet, conditionnellement à la date de rupture –significative- identifiée par le test SupF(1), la présence d'une seconde rupture additionnelle n'est significative dans aucun des cas. L'approche séquentielle préconisée par Bai et Perron (1998b) conduit donc à conclure à la présence d'une seule rupture sur la période 1963-2000, située au voisinage de 1974.

Pour évaluer la robustesse de ce résultat, *i.e.* l'unicité de la rupture, nous estimons les modèles sur la période 1974 :1-2001 :1 et testons la présence de rupture au sein de cette période (tableau 4). Les différents tests effectués sur les modèles autorégressifs (*i.e.* ceux relevant de l'approche paramétrique, reportés dans la partie inférieure du tableau) concluent à la stabilité de la tendance sur cette nouvelle période. Dans le cas de l'approche non-paramétrique, la stabilité ne peut également être rejetée pour le modèle stationnaire en différence. Pour ce qui est du modèle TS, le résultat est moins net : le test ExpF conduit à retenir l'hypothèse de stabilité mais le test AvgF rejette la stabilité à 1%. Le test SupF(1) rejette la stabilité au seuil de 10% (indiquant une rupture en 1979 :1) mais non au seuil de 5 %. Le test SupF(2) mené sur le modèle TS suggère de son côté une double rupture en 1991:3 et 1996:4 significative au seuil de 1%. La première rupture (à la baisse) pourrait s'interpréter comme l'effet d'un ralentissement du progrès technique ou l'effet sur la productivité des mesures favorables à l'emploi non qualifié⁽¹⁵⁾, et la deuxième comme l'effet sur la productivité des nouvelles technologies et de la « nouvelle économie ». Toutefois la présence de ces deux ruptures reste statistiquement fragile. D'une part, si le test mené sur le modèle DS confirme cette datation des ruptures (l'une au début des années 1990, l'autre vers 1997), elles n'apparaissent pas significatives dans ce modèle. D'autre part, si l'on suit une démarche de test séquentielle, on est conduit à ne retenir aucune rupture à la suite du test $m=0$ contre $m=1$ rupture pour onze des douze tests réalisés (au seuil de 5%). Enfin, les simulations menées dans l'annexe 2 indiquent que le test fondé sur le modèle TS associé à la correction non-paramétrique tend, en échantillon fini, à détecter trop souvent des ruptures, même lorsque l'hypothèse de stabilité est en fait vérifiée (en particulier si le vrai modèle est en fait DS). Cette propriété nous conduit donc à accorder un moindre crédit aux cas de rejet pour le test réalisé pour cette spécification et avec cette procédure. Le diagnostic obtenu pour la première rupture (tableau 2) nous paraît néanmoins rester fiable, car confirmé par les trois autres procédures de tests, qui ont des performances (en terme de taille) beaucoup plus satisfaisantes.

Au total, les tests effectués conduisent ainsi à retenir comme la plus pertinente une modélisation avec une unique rupture, en 1973 ou 1974, la stabilité sur la période postérieure à 1973 ne pouvant être rejetée. Afin d'obtenir des

Tableau 4 : tests de stabilité contre des ruptures multiples, période 1974 :1-2001:4

	Modèle TS	Modèle DS
Approche non paramétrique	eq. (15)	eq. (17)
Test de $m=0$ vs. $m=1$ rupture		
ExpF	1,41	0,25
AvgF	2,81**	0,43
SupF	8,64*	2,04
Dates	1979:1	1997:1
Test de $m=0$ vs. $m=2$ ruptures		
Statistique SupF(2)	37,96***	5,55
Dates	1991:3, 1996:4	1990:1, 1997:1
Approche paramétrique	eq. (16)	eq. (18)
Test de $m=0$ vs. $m=1$ rupture		
ExpF	0,36	0,207
AvgF	0,62	0,37
SupF	2,54	1,62
Dates	1977:1	1990:1
Test de $m=0$ vs. $m=2$ ruptures		
SupF(2)	3,15	4,23
Dates	1991:3, 1996:4	1987:1, 1990:1

Note : période : 1963:1-2001:4 ; */**/** : respectivement significatif à 10%, 5%, 1% ; Les valeurs critiques utilisées sont données dans les tableaux 2 et 3. Dates : date(s) de rupture présumée(s).

Tableau 5 : intervalles de confiance

		Date	Intervalle de confiance à 90%
Modèle TS	Approche non paramétrique	1973:4	[1973 :1-1974 :2]
Modèle DS	Par approche paramétrique	1974:1	[1968:1-1975 :3]
	Approche non paramétrique	1974:1	[1972:4-1976 :3]

intervalles de confiance autour de cette date de rupture, nous appliquons la procédure de Bai (1997) aux équations (17) et (18), et celle de Perron et Zhu (2002) à l'équation (15). Les résultats sont présentés dans le tableau 5. Les intervalles à 90% obtenus sont relativement étroits. L'intervalle de confiance le plus large est obtenu pour le modèle D avec approche paramétrique : il va du premier trimestre 1968 au troisième trimestre 1975. L'année 1973 et les deux premiers trimestres de 1974 sont inclus dans tous les intervalles de confiance. On en conclut que les différences entre les dates de rupture concurrentes obtenues par les différentes procédures ne sont pas significatives.

Implications pour la croissance tendancielle

La datation des ruptures obtenue peut être utilisée pour évaluer rétrospectivement la croissance tendancielle. Pour cela le tableau 6 présente les estimations des quatre spécifications (15) à (18) testées, estimées dans chaque cas pour la date de rupture optimale. Les équations ont des qualités d'ajustement très différentes, car dans l'approche paramétrique la dynamique est explicitement modélisée tandis qu'elle est incorporée au résidu dans le cas des deux approches non paramétriques (un test LM illustre bien que l'absence d'autocorrélation des résidus est rejetée pour les deux spécifications paramétriques tandis qu'elle n'est pas rejetée pour les modèles autorégressifs). Les quatre spécifications fournissent en revanche des résultats très proches pour l'évaluation de la croissance tendancielle. Ce résultat est dû à la proximité des dates de ruptures obtenues (et au fait que, sur une période donnée, la moyenne empirique de Δy_t est proche de l'estimateur \hat{g} de g dans le modèle $y_t = a + gt + e_t$). La croissance tendancielle est évaluée à 5 % environ (4,9 à 5,1 selon les cas) avant 1974 et 2,1 % après 1974. L'évaluation de la croissance tendancielle sur la période la plus proche est donc relativement pessimiste en comparaison avec des évaluations structurelles de la croissance potentielle formulées plus récemment (voir par exemple De Bandt et Villette, 2002). Si l'évaluation de la croissance tendancielle de moyen-long terme est proche selon les modèles, calculer un écart de production de court terme associé à chacun des modèles nécessite des hypothèses supplémentaires et aboutirait à des profils vraisemblablement contrastés. À titre illustratif, l'annexe 3 présente un calcul d'écart de production fondé sur le modèle TS statique, correspondant à l'approche traditionnelle par la tendance segmentée.

Tableau 6 : estimation des quatre spécifications de référence pour les dates de rupture optimales

Modèle TS-NW (équation 15)			Modèle TS-AR (équation 16)			Modèle DS-NW (équation 17)			Modèle DS-AR (équation 18)		
Date de rupture t*	1973:4		1972:4			1973:4			1974:1		
Param.	Coeff.	t-stat	Param.	Coeff.	t-stat	Param.	Coeff.	t-stat	Param.	Coeff.	t-stat
μ	4638,82	3248	μ	421,57	3,79	β_1	5,11	13,88	β_1	5,12	13,71
β_1	5,11	73,7	β_1	4,86	15,72	$\beta_2 (t > t^*)$	-2,97	-6,43	$\beta_2 (t > t^*)$	-2,98	-6,75
$\beta_2 (t > t^*)$	-3,01	-31	$\beta_2 (t > t^*)$	-2,80	-8,82				c1	0,12	1,57
			c1	1,07	13,39				c2	0,06	0,76
			c2	-0,05	-0,48				c3	0,19	2,37
			c3	0,14	1,20				c4	-0,26	-3,28
			c4	-0,44	-3,84						
			c5	0,19	2,50						
LM(4)	0,00		LM(4)	0,97		LM(4)	0,00		LM(4)	0,45	
ser	6,55		ser	2,17		ser	2,31		ser	2,21	
Croissance tendancielle											
1963:1-1973:4	5,1		1963:1-1972:4	4,9		1963:1-1973:4	5,1		1963:1-1974:1	5,1	
1974:1-2001:4	2,1		1973:1-2001:4	2,1		1974:1-2001:4	2,1		1974:2-2001:4	2,1	

Note : le logarithme du PIB est multiplié par 400 pour permettre l'interprétation des paramètres en termes de points de croissance annuelle ; SER: écart-type de l'estimation ; t-stat : T de Student. Pour les spécifications TS-AR et DS-AR, les modèles sont estimés par moindres carrés non-linéaires (afin d'estimer directement la croissance de long terme). Pour les spécifications TS-NW et DS-NW les t-stats sont corrigés de l'autocorrélation-hétéroscédasticité selon la méthode de Newey-West (1987) décrite dans l'équation (11) (avec un générateur de fenêtre de Bartlett et une fenêtre de 4 trimestres). LM(4) : p-value d'un test LM(4) d'autocorrélation des résidus.

Conclusion

Cet article avait pour objectif de décrire des méthodes relativement récentes d'analyse économétrique des changements structurels, et d'en proposer une application à la croissance tendancielle du PIB tendanciel français.

En ce qui concerne nos résultats empiriques, l'application de tests récents de changements structurels confirme l'existence d'une rupture significative dans la croissance tendancielle française vers 1973, au moment du premier choc pétrolier. Aucune rupture significative n'est détectée sur la période postérieure à 1974. Ces résultats sont robustes à l'hypothèse faite sur la non-stationnarité du PIB, *i.e.* la stationnarité autour d'une tendance segmentée ou une marche aléatoire avec dérive. Ils invitent donc à considérer avec prudence l'hypothèse parfois envisagée d'un ralentissement de la production tendancielle au début des années 1990, comme celle d'une accélération récente et durable de la croissance tendancielle liée à la « nouvelle économie ». Un diagnostic plus précis ne sera bien sûr possible qu'avec le recul fourni par plusieurs années d'observations supplémentaires. Toutefois, une voie de recherche pour affiner ce diagnostic pourrait être de combiner les méthodes ici utilisées avec une approche structurelle de la croissance potentielle.

Il convient à cet égard de souligner que les utilisations possibles du type de tests présentés ne se limitent pas au cas assez simple retenu dans l'application. Par exemple, les tests de stabilité en question ont été employés pour examiner l'hypothèse de l'atténuation de l'ampleur des fluctuations cycliques aux États-Unis⁽¹⁶⁾, ou pour évaluer la pertinence de la critique de Lucas dans des modèles macroéconomiques. La critique de Lucas prédit en effet que les paramètres d'une forme réduite décrivant l'économie devraient se révéler instables lors d'une modification de la règle de politique économique, car ces paramètres dépendent, *via* les anticipations rationnelles des agents, de ceux régissant la politique économique. Des tests de stabilité utilisés pour détecter si une forme réduite connaît une rupture simultanément ou postérieurement à un changement de règle de politique économique, peuvent ainsi être interprétés comme des tests de la critique de Lucas⁽¹⁷⁾.

L'extension récente des tests de stabilité à un cadre multivarié devrait élargir encore leur champ d'application. Bai, Lumsdaine et Stock (1998) ont ainsi proposé une procédure d'inférence sur la date de rupture dans un cadre multivarié -dans le cas d'une rupture unique-, et montré que l'estimation des ruptures dans ce contexte permet un gain important de précision dans la datation des ruptures.

Notes

(1) À notre connaissance, peu de travaux ont à ce jour appliqué ces méthodes à la question du PIB tendanciel. L'étude de Filardo et Cooper (1996) pour les États-Unis constitue une exception.

(2) Pour des discussions plus détaillées de ces notions et une comparaison des approches, nous renvoyons par exemple à Boccara, Bouthevillain, Cœure et Eyssartier (1997), Le Bihan, Sterdyniak et Cour (1997) ou aux contributions recueillies dans Baghli *et alii* (2002).

(3) En revanche, pour d'autres objectifs comme la prévision, l'approche par MCO statique serait clairement inadaptée en cas d'une significative autocorrélation des résidus.

(4) Soulignons que dans ce contexte, un modèle DS sans rupture est testé contre un modèle TS avec rupture.

(5) Hamilton (1994, p.652) recommande une approche analogue, la qualifiant d' "éclectique".

(6) En effet pour un échantillon donné, la somme des carrés sous l'hypothèse nulle $SSR_{1,T}$ est constante, et l'équation (9) indique donc que la statistique F est une fonction décroissante de la somme des carrés des résidus $SSR_{A_{T_1}}$ obtenue en autorisant une rupture.

(7) Les propriétés théoriques (convergence et distribution des estimateurs des dates de rupture) portent en fait sur la fraction T_1 / T . Pour un échantillon donné, il est cependant équivalent de raisonner en termes de date de rupture.

(8) Nous renvoyons au document de travail de Bai et Perron (1998b) plutôt qu'à la version publiée (Bai et Perron, 2003) car il contient notamment la tabulation des valeurs critiques des tests.

(9) Liu, Wu et Zidek (1997) proposent un critère d'information alternatif au BIC :

$$LWZ = \ln(SSR / (T - p^*)) + (p^* / T) c_0 [\ln(T)]^{(2 + \delta_0)}$$

avec $p^* = (m + 1)q + m + p$, et suggèrent $c_0 = 0,299$ et $\delta_0 = 0$.

(10) Nous remercions S.Gregoir (Insee) pour nous avoir communiqué cette série longue.

(11) L'examen visuel de la série indique que le second trimestre est marqué par une valeur extrême, mais que ce choc est complètement transitoire. Nous suivons l'interpolation utilisée par Banerjee *et alii* (1992) et Vogelsang (1997) pour nous prémunir contre un biais lié à ce "point aberrant". Le second auteur indique que ses résultats sont en effet affectés par l'outlier de 1968:2.

(12) Cette valeur de K_{\max} coïncide avec la suggestion d'Hayashi (2000, p.594) de prendre K_{\max} égal à la partie entière de $12(T/100)^{1/4}$

(13) Nous avons vérifié que les résultats des tests de stabilité présentés *infra* sont robustes au choix de l'hypothèse $p_{\max} = 13$.

(14) Ces seuils sont approximatifs pour deux raisons. D'une part, ils correspondent à des valeurs critiques asymptotiques. Simuler des valeurs critiques exactes à distance finie dépasse le cadre de notre exercice, notamment en raison de l'important volume de calcul requis dans le cas de ruptures multiples (pour chaque réplcation de Monte Carlo, l'estimation devrait être menée pour chacun des "m-uplets" de dates de ruptures possibles). D'autre part, les articles cités étudient des modèles caractérisés par des ruptures sur des paramètres individuels, tandis que le modèle de changement de pente utilisé dans le cas TS se distingue en introduisant une contrainte sur les paramètres lors de la rupture.

(15) Voir Duchêne, Forgeot, Jacquot (1997) pour une discussion de l'hypothèse d'un ralentissement de la productivité tendancielle au début des années 90, et de ses causes possibles.

(16) Voir notamment Mc Connell et Perez-Quiros (2000), Stock et Watson (2003).

(17) Voir par exemple Estrella et Fuhrer (1999), Hénin et Podevin (2002) ou Jondeau et Le Bihan (2002).

Bibliographie

- Andrews D.K. (1993).** « Tests for Parameter Instability and Structural Change With Unknown Changepoint », *Econometrica*, vol. 61, 4, pp. 821-856.
- Andrews D.K. et Ploberger W. (1994).** « Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative », *Econometrica*, vol. 62, 6, pp. 1383-1414.
- Andrews D.K., Lee I. et Ploberger W. (1996).** « Optimal Change Point Tests for Normal Linear Regression », *Journal of Econometrics*, vol. 70, pp. 9-38.
- Bai J. (1997).** « Estimation of a Changepoint in Multiple Regression Models », *Review of Economics and Statistics*, vol. 79, pp. 551-563.
- Bai J. (1999).** « Likelihood Ratio Tests for Multiple Structural Changes », *Journal of Econometrics*, vol. 91, pp. 299-323.
- Bai J., Lumsdaine R. et Stock J. (1998).** « Testing for and Dating Common Breaks In Multivariate Time Series », *Review of Economic Studies*, vol. 65, 3, pp. 375-432.
- Bai J. et Perron P. (1998 a).** « Estimating and Testing Linear Models with Multiple Structural Breaks », *Econometrica*, vol. 66, 1, pp. 47-78.
- Bai J. et Perron P. (1998 b).** « Computation and Analysis of Structural Change Models », Cahier 0998, CRDE, *Université de Montréal*.
- Bai J. et Perron P. (2003).** « Computation and Analysis of Structural Change Models », *Journal of Applied Econometrics*, vol. 18, pp. 1-22.
- Banerjee A., Lumsdaine R. et Stock J. (1992).** « Recursive and Sequential Tests of the Unit-Root and trend-Break Hypotheses: Theory and international Evidence », *Journal of Business and Economic Statistics*, vol. 10, 3, pp. 271-287.
- Baghli M., Bouthevillain C., De Bandt O., Fraisse H., Le Bihan H. et Rousseaux P. (2002).** « PIB potentiel et écart de PIB. Quelques évaluations pour la France », *Note d'Etudes et de Recherche* #89, Banque de France.
- Boccara F., Bouthevillain K., Cœure B. et Eyssartier D. (1997).** « Comment positionner les économies dans le cycle », *Economie Internationale*, n° 69, pp. 55-76.
- Canjels E. et Watson M. (1997).** « Estimating Deterministic Trends in the Presence of serially Correlated Errors », *Review of Economics and Statistics*, vol. 79, 2, pp. 184-200.
- Canova F. (1998).** « Detrending and Business Cycle Facts », *Journal of Monetary Economics*, vol. 41, pp. 475-512.
- Chow G.C. (1960).** « Tests of Equality Between Sets of Coefficients in Two Linear Models », *Econometrica*, vol. 28, pp. 591-605.
- De Bandt O. et Villetelle J.P. (2002).** « Analyse du positionnement dans le cycle par les indicateurs de croissance potentielle et d'écart de production quelques évaluations pour la France », *Bulletin de la Banque de France*, n° 103, pp. 55-67.
- Doz C., Rabault G. et Sobczak N. (1995).** « Décomposition tendance-cycle : estimation par des méthodes statistiques univariées », *Économie et Prévision*, n° 120, vol. 4, pp. 73-93.
- Duchêne S., Forgeot G. et Jacquot A. (1997).** « Les évolutions récentes de la productivité du travail », *Économie et Statistique*, n° 301-302, pp. 169-192.
- Estrella R. et Fuhrer J. (2003).** « Monetary Policy Shifts and the Stability of Monetary Policy Models », *Review of Economics and Statistics*, vol. 85, 1, pp. 94-104.
- Filardo A. et Cooper P. (1996).** « Cyclically- Adjusted Measures of Structural Trend Breaks: an Application to Structural Trend Breaks in the 1990s », *Research Working Paper* 96-14, Federal Reserve Bank of Kansas City.
- Hamilton J. (1994).** *Time Series Analysis*, Princeton University Press, Princeton.
- Hansen B. (2001).** « The New Econometrics of Structural Changes: Dating Breaks in Us Labor productivity », *Journal of Economic Perspectives*, vol. 15, 4, pp. 117-128.
- Hayashi F. (2000).** *Econometrics*, Princeton University Press, Princeton.
- Hénin P.Y. et Podevin M. (2002).** « Assessing the Effects of Policy Change : Lessons from the European 1992 experience », *Annales d'Economie et de Statistique*, n° 67-68, pp. 435-461.
- Jondeau E. et Le Bihan H. (2002).** « Evaluating Monetary Policy Rules in Estimated Forward-Looking Models. A Comparison of US and German Monetary Policies », *Annales d'Economie et de Statistique*, n° 67-68, pp. 357-381.
- Kim I.M. (1997).** « Detecting the Number of Structural Breaks », *Economic Letters*, vol. 57, pp. 145-148.
- Le Bihan H., Sterdyniak H. et Cour P. (1997).** « La notion de croissance potentielle a-t-elle un sens ? », *Economie Internationale*, n° 69, pp. 17-53.
- Liu J., Wu S. et Zidek J.V. (1997).** « On Segmented Multivariate Regressions », *Statistica Sinica*, vol. 7, pp. 497-525.

- Lumsdaine R et Papell D. (1997).** « Multiple Trend Breaks and the Unit-Root Hypothesis », *Review of Economics and Statistics*, vol. 79, pp. 212-218.
- Mc Connell M. et Perez Quiros G. (2002).** «Output Fluctuations in the United States : What Has Changed Since the Early 1980's ? », *American Economic Review*, vol. 90, n° 5, pp. 1464-1476.
- Newey W. et West K. (1987).** « A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix », *Econometrica*, vol. 55 , n° 3, pp. 703-708.
- Perron P. (1997).** « Further Evidence on Breaking Trend Functions in Macroeconomic Variables », *Journal of Econometrics*, vol. 80, pp. 355-385.
- Perron P. et Zhu X. (2002).** « Structural Breaks with Deterministic and Stochastic Trends », *mimeo*, Boston University.
- Stock J. (1994).** « Unit Roots, Structural Breaks and Trends » Chap.46 in *Handbook of Econometrics*, vol. IV, Ed. R.F.Engle et D.L. Mc Fadden.
- Stock J. et Watson M. (2003).** « Has the Business Cycle Changed and Why? », *NBER Macroeconomics Annual 2002*, vol. 17, pp. 159-218.
- Yao Y.C. (1988).** «Estimating the Number of Change Points Via Schwartz' Criterion », *Statistics and Probability Letters*, vol. 6, pp. 181-189.
- Zivot E. et Andrews D.K. (1992).** « Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis », *Journal of Business and Economic Statistics*, vol. 10, 3, pp. 251-270.

Annexe 1 : robustesse de la détection des dates de rupture à la spécification de la tendance

Pour illustrer la robustesse de la détection des dates de rupture à la spécification de la tendance nous menons l'exercice de Monte-Carlo suivant.

Le modèle simulé est

$$(A1) \Delta y_t = \mu_1 I(t \leq T_1) + \mu_2 I(t > T_1) + v_t$$

où v_t est un bruit blanc. Ce modèle est analogue au modèle DS du texte : la variable y_t est alors non-stationnaire intégrée, et la composante déterministe de son taux de croissance connaît de plus une rupture à la date T_1 .

Les paramètres retenus sont $\mu_1 = 5,0$, $\mu_2 = 2,5$ afin de répliquer la baisse du taux de croissance moyen observée en France au voisinage du premier choc pétrolier. L'aléa v_t suit une loi normale $N(0,1)$. Nous choisissons $T = 200$ et une rupture située à la date $T_1 = T / 2$.

Pour chaque simulation, une trajectoire de taille T est simulée et deux modèles sont estimés à partir de la trajectoire simulée, conduisant à deux statistiques de test :

– le modèle DS (D), bien spécifié, est estimé successivement pour chacune des dates de rupture possibles entre l'observation 10 et l'observation $T - 10$. La statistique SupF est calculée, ce qui fournit un estimateur \hat{T}_1^{DS} de la date de rupture (supposée inconnue de l'économètre) ;

– le modèle TS (B), mal spécifié, est également estimé pour toutes les dates de rupture possibles, et la statistique SupF fournit alors un second estimateur \hat{T}_1^{TS} de la date de rupture.

On réalise $N=2000$ simulations et on peut étudier la distribution des deux estimateurs \hat{T}_1^{DS} et \hat{T}_1^{TS} . Cette distribution est représentée par un histogramme dans le graphique A1 et est résumée dans le tableau A1.

On observe que les deux modèles fournissent des estimateurs sans biais de la vraie date de rupture. On a par ailleurs vérifié (en menant un exercice de Monte Carlo avec $T=10000$) que les deux estimateurs étaient convergents. D'autre part, les deux estimateurs ont une précision comparable, en termes d'intervalles interquartiles. Le modèle bien spécifié détecte plus souvent la vraie valeur (8,7% contre 4,4% pour le modèle *trend*-stationnaire) mais présente plus souvent des valeurs « extrêmes », ce qui conduit à un écart type plus élevé.

Graphique A1 : distribution des estimateurs \hat{T}_1

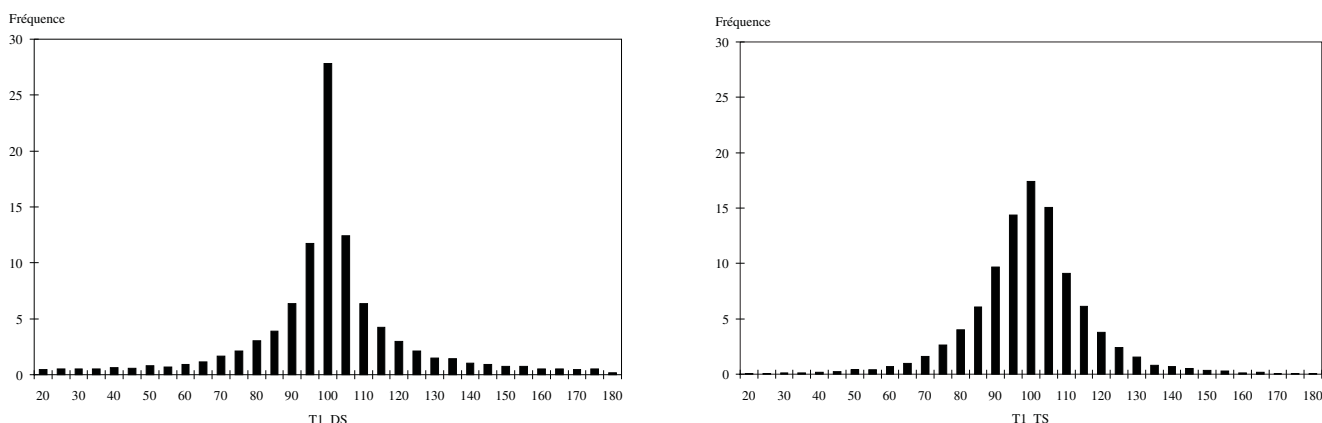


Tableau A.1: distribution de l'estimateur de la date de rupture

Estimateur	\hat{T}_1^{DS}	\hat{T}_1^{TS}
Médiane	100	100
Écart type	22,3	16,1
Intervalle inter quantile (25% -75%)	[93 – 107]	[92 – 108]
Pr ($T_1 = 100$)	8,70%	4,40%

Annexe 2 : impact du mode de prise en compte de l'autocorrélation, et de la spécification de la non-stationnarité sur les tests de stabilité

Nous explorons ici les implications du mode de correction de l'autocorrélation, et du choix de la spécification estimée, en simulant des modèles DS et TS incorporant une autocorrélation des résidus. Des tests de rupture sont alors menés successivement en utilisant une correction paramétrique et non paramétrique de l'autocorrélation des résidus, et pour une spécification estimée DS et TS. L'objectif est d'éclairer certains de nos résultats empiriques, une étude plus complète sur les propriétés à distance finie des tests employés dépassant le cadre de cet article.

Les deux types de DGP (*data generating process*) simulés sont un modèle DS :

$$(A2) \quad \Delta y_t = \mu_1 I(t \leq T_1) + \mu_2 I(t > T_1) + v_t$$

$$(A3) \quad v_t = \rho_e v_{t-1} + e_t$$

et un modèle TS :

$$(A4) \quad y_t = \alpha + \mu_1 t I(t \leq T_1) + \mu_2 (t - T_1) I(t > T_1) + u_t$$

$$(A5) \quad u_t = \rho u_{t-1} + \varepsilon_t$$

où e_t et ε_t sont des bruits blancs suivant des lois normales $N(0, \sigma_e^2)$ et $N(0, \sigma^2)$. On retient $T = 200$ et une rupture située à la date $T_1 = T/2$. Les paramètres sont choisis pour approcher ceux estimés sur données réelles : on considère $\mu_1 = 5,0$, $\mu_2 = 2,5$ ainsi que $\mu_1 = 2,5$ (dans ce dernier cas pour étudier la taille du test). On considère également $\rho_e = 0,3$ ou $0,8$, $\rho = 0,9$ (notons que dans le cas $\rho = 0,9$ la « demi-vie » d'une innovation sur le PIB vaut $\ln(0,5)/\ln(\rho) = 6,58$ trimestres), $\sigma_e = 2$ ou 5 , $\sigma = 2$ ou 10 (les résultats avec $\sigma = 5$ ne différant pas du cas $\sigma = 2$) et, sans restriction, $\alpha = 0$. Au total, dix DGP illustratifs sont simulés, correspondant aux dix jeux de paramètres renseignés dans la première colonne du tableau A2.

Pour chaque simulation, une trajectoire de taille T est simulée et quatre tests de stabilité sont réalisés, correspondant aux quatre spécifications suivantes :

- une équation DS de type (A2) estimée par moindres carrés ordinaires, notée DS_NW dans le tableau A2 ;
- une équation DS autorégressive :

$$\Delta y_t = \mu_1 I(t \leq T_1) + \mu_2 I(t > T_1) + \sum_{i=1}^p c_i \Delta y_{t-i} + u_t$$

Tableau A.2 : puissance et taille à distance finie du test de rupture SupF

					Modèle estimé :			
DGP: DS	μ_1	μ_2	ρ_e	σ_e	DS_AR	DS-NW	TS_AR	TS_NW
(1)	5,00	2,50	0,30	2	0,99	1,00	0,94	1,00
					<i>99,70</i>	<i>100,10</i>	<i>91,70</i>	<i>100,20</i>
(2)	5,00	2,50	0,30	5	0,50	0,61	0,62	0,99
					<i>100,30</i>	<i>100,70</i>	<i>95,30</i>	<i>100,10</i>
(3)	5,00	2,50	0,80	2	0,36	0,77	0,60	0,97
					<i>98,40</i>	<i>99,70</i>	<i>98,10</i>	<i>100,10</i>
(4)	5,00	2,50	0,80	5	0,14	0,52	0,37	0,94
					<i>98,30</i>	<i>99,30</i>	<i>102,60</i>	<i>99,10</i>
(5)	2,50	2,50	0,30	2	0,07	0,13	0,29	0,91
					<i>98,80</i>	<i>98,10</i>	<i>102,30</i>	<i>98,80</i>
(6)	2,50	2,50	0,30	5	0,05	0,12	0,28	0,92
					<i>97,80</i>	<i>98,00</i>	<i>102,90</i>	<i>98,60</i>
DGP: TS	μ_1	μ_2	ρ_e	σ				
(7)	5,00	2,50	0,90	2	1,00	1,00	1,00	1,00
					<i>100,30</i>	<i>100,00</i>	<i>96,00</i>	<i>100,10</i>
(8)	5,00	2,50	0,90	10	0,12	0,26	0,88	1,00
					<i>99,90</i>	<i>99,30</i>	<i>96,00</i>	<i>100,40</i>
(9)	2,50	2,50	0,90	2	0,00	0,00	0,08	0,43
					<i>99,90</i>	<i>98,40</i>	<i>100,10</i>	<i>97,60</i>
(10)	2,50	2,50	0,90	10	0,00	0,01	0,08	0,42
					<i>101,40</i>	<i>99,60</i>	<i>103,30</i>	<i>100,20</i>

Note : taux de rejet de l'hypothèse nulle et (en italiques) date de rupture moyenne estimée .

avec $p = 4$. Cette spécification est notée DS_AR dans le tableau.

- une équation TS de type (A4). La spécification est dénommée TS_NW ;
- une équation TS autorégressive, notée TS_AR

Pour chaque cas, un test de stabilité est réalisé de la façon suivante. Une date de rupture optimale est déterminée en minimisant la somme des carrés des résidus. Pour cette date, dans le cas des spécifications DS_AR et TS_AR, la statistique SupF est calculée. Dans le cas de DS_NW et TS_NW, la statistique SupF* robuste à l'autocorrélation et l'hétéroscédasticité des résidus est calculée en utilisant la procédure de Newey-West (1987) avec un paramètre de fenêtre $L = 4$. Si la statistique SupF (ou SupF*) dépasse la valeur critique à 5% tabulée dans Andrews (1993) (ici 8,85 pour 1 paramètre et $((T_{\min} / T) = 0,15)$ on rejette la stabilité.

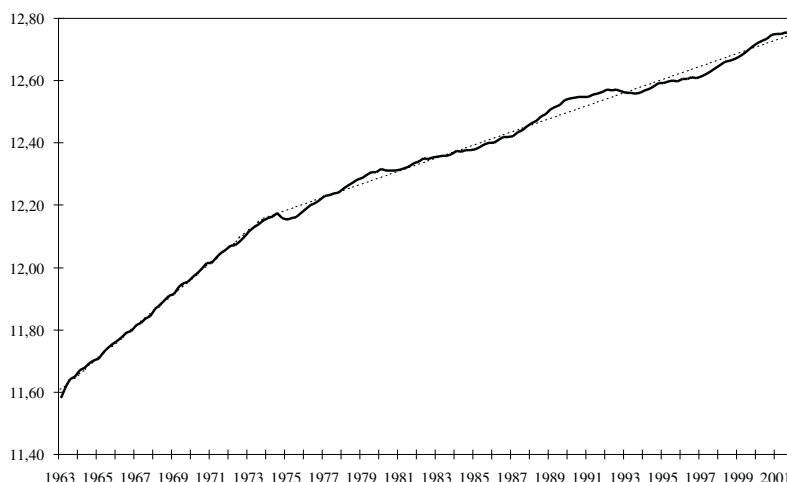
On réalise $N=1000$ simulations. Pour chaque DGP, le tableau A2 donne la proportion de cas (arrondie à deux décimales) pour lesquels l'hypothèse de stabilité est rejetée (en italique figure la moyenne des estimateurs de la date de rupture obtenus). Le taux de rejet de l'hypothèse nulle fournit une indication sur la puissance (la probabilité de rejeter l'hypothèse nulle si elle est fausse) à distance finie des différentes procédures de test lorsque $\mu_1 = \mu_2$, et sur leur taille à distance finie lorsque $\mu_1 \neq \mu_2$ (i.e pour les DGP 5,6,9 et 10).

Quelques-uns des enseignements de l'exercice sont les suivants. Tout abord, toutes les procédures présentent une certaine capacité à détecter la rupture, même lorsque la spécification de la non-stationnarité employée est erronée. La puissance de chacun des tests diminue quand la variabilité ou la persistance des résidus s'accroît (ce qu'indique la comparaison des lignes (2) et (1) d'une part, (3) et (1) d'autre part). Le changement structurel devient dans ces cas masqué par l'ampleur des fluctuations transitoires. Par ailleurs, l'utilisation d'une correction non-paramétrique conduit systématiquement à rejeter plus souvent l'hypothèse nulle : le taux de rejet affiché dans la colonne DS_NW (respectivement TS_NW) est supérieur à celui affiché dans la colonne DS_AR (respectivement TS_AR). Ceci est cohérent avec les résultats obtenus en échantillon. Pour les DGP simulés, la procédure non-paramétrique est plus puissante pour détecter la rupture. Toutefois cet avantage en terme de puissance a un coût important en terme de niveau du test (i.e. dans les cas $\mu_1 = \mu_2 = 2,5$) : la correction non-paramétrique tend à rejeter plus souvent l'hypothèse nulle -à tort dans ce cas. On constate également que le test associé au modèle TS conduit à rejeter plus souvent la stabilité que celui associé au modèle DS. Plus particulièrement, les performances du test associé au modèle TS avec correction non paramétrique sont peu satisfaisantes (dans le DGP (9) pour lequel l'estimateur TS_NW est bien spécifié, le test rejette à tort la stabilité dans 42% des cas à distance finie).

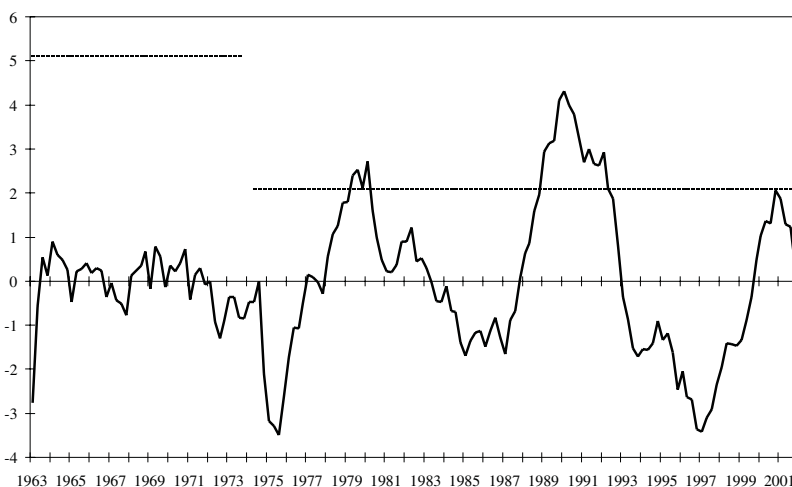
Annexe 3 : évaluation d'un écart de production sous l'hypothèse TS

En retenant l'hypothèse d'un modèle TS avec rupture estimé par MCO « statiques », on peut déduire très simplement du modèle estimé un écart de production implicite : il correspond au résidu de l'équation estimée. Pour cet exercice, le choix de l'hypothèse TS plutôt que DS est crucial (et sujet à controverse). Soulignons que le modèle TS n'est pas nécessairement le meilleur modèle du PIB, et que l'exercice est ici réalisé à titre illustratif. Le PIB tendanciel et l'écart de production associés au modèle à une rupture sont présentés dans les graphiques A2 et A3. Au cours du dernier cycle, l'écart de production atteindrait un point haut en 1990:1 (4,3 %) et un point bas en 1997:1 (-3,4 %). L'année 2000 marquerait un nouveau point haut. Par rapport aux autres évaluations (obtenues par des filtres Hodrick-Prescott ou des modèles à composantes inobservables), les fluctuations sont de plus grande ampleur. Cette grande variabilité de la composante cyclique obtenue par tendance linéaire est habituelle : nous renvoyons par exemple à Canova (1998) pour la comparaison des propriétés des composantes cycliques obtenues par différentes méthodes, et à De Bandt et Villette (2002) pour des résultats sur la France.

Graphique A2 : PIB et PIB tendanciel en logarithme



Graphique A3 : écart de production et croissance tendancielle



Note : en % ; croissance tendancielle (annualisée) en pointillés.