

Modèles VAR ou DSGE: Que choisir ?*

Fabrice Collard

Université de Toulouse 1 (CNRS–GREMAQ et IDEI)

Patrick Fève[†]

Université de Toulouse 1 (GREMAQ et IDEI)

Banque de France (Division de la Recherche)

Septembre, 2006

Résumé

L'objet de cet article est de comparer les performances relatives des modèles VAR et DSGE. Les modèles sont estimés par maximum de vraisemblance sur le taux de croissance de la productivité horaire du travail et les heures travaillées aux Etats-Unis pour la période 1959–2003. Nous montrons que les modèles DSGE, lorsqu'ils intègrent suffisamment de rigidités réelles (coûts d'ajustement sur l'investissement ou persistance des habitudes de loisir) ont des qualités prédictives au moins égales à celles d'un modèle VAR. Ces résultats empiriques suggèrent que, de par leur caractère structurel, les modèles DSGE doivent être préférés aux modèles VAR, notamment lorsqu'il s'agit d'identifier les chocs structurels et leurs effets sur la dynamique agrégée.

Mots-clés : Modèles VAR, Modèles DSGE, Méthodes d'évaluation quantitatives, Identification

Classification JEL : C3, C5, E3

*Nous tenons à remercier J.P. Laffargue ainsi qu'un rapporteur pour leur nombreuses remarques et suggestions. Nous demeurons seuls responsables d'éventuelles erreurs et omissions. Celles-ci n'engagent en aucun cas la Banque de France.

[†]Adresse: GREMAQ–Université de Toulouse 1, manufacture des Tabacs, bât. F, 21 allée de Brienne, 31000 Toulouse. email: patrick.feve@univ-tlse1.fr.

1 Introduction

L'analyse quantitative des fluctuations agrégées s'est, dans une large mesure, développée autour de modèles macro-économétriques de relativement grande taille. Ces modèles sont, par exemple, utilisés par les administrations ou les institutions publiques à des fins d'évaluation quantitative de différentes options possibles pour la conduite de la politique économique, mais aussi à des fins de prévision économique.

Jusqu'au début des années soixante-dix, il existait un relatif consensus sur le modèle théorique de référence — le modèle Offre Globale/Demande Globale en présence de certaines rigidités nominales —, la méthodologie statistique d'évaluation et son utilisation. La modélisation macro-économétrique ne donnait alors lieu qu'à peu de controverses.

A la suite des critiques de Sims (1980) et de Lucas (1976), deux autres types de modélisation se sont progressivement imposées, au moins à un niveau académique : les modèles Vectoriels Auto-Régressifs (VAR) et les Modèles d'Équilibre Général Intertemporels Stochastiques (DSGE). Ces deux approches ont développé et proposé des outils quantitatifs originaux permettant de renouveler l'analyse des fluctuations. D'un côté, l'approche VAR permet de modéliser la dynamique de différentes variables agrégées à l'aide d'un faible nombre de restrictions, la sélection du modèle ne s'effectuant que sur la base de critères statistiques. D'un autre côté, les modèles DSGE offrent une modélisation parcimonieuse et rigoureuse de la dynamique économique et des anticipations.

Jusqu'à très récemment, les modèles VAR et DSGE sont restés largement spécialisés, chacun ayant un rôle et donnant lieu à une utilisation spécifique. Les modèles VAR sont des outils statistiques et descriptifs qui fournissent un très bon ajustement aux données. Il est alors possible de réaliser facilement des exercices de prévision à différents horizons des variables macroéconomiques à partir de la représentation récursive du modèle VAR sans avoir à formuler des hypothèses supplémentaires sur l'environnement économique. Les modèles DSGE sont, quant à eux, utilisés pour mener différents exercices théoriques et appliqués dans un cadre immune à la *critique de Lucas*.

Cependant, ce partage des rôles a été progressivement remis en cause. Ainsi le caractère a-théorique de la méthodologie VAR a rapidement laissé place à une interprétation structurelle des résultats. En effet, les modèles VAR permettent, moyennant l'introduction d'un nombre minimal de restrictions, d'identifier certains chocs structurels (voir Sims (1980), Sims (1986), Christiano, Eichenbaum et Evans (1999), Blanchard et Quah (1989), Galí (1999), Christiano, Eichenbaum et Evans (2005)) et d'en dériver les effets sur les variables agrégées. En conséquence, les modèles VAR ont d'abord pris un ascendant considérable

sur les modèles DSGE dans la mesure où ils ont été perçus comme des guides extrêmement utiles pour la modélisation théorique.¹ Ainsi, les fonctions de réponses des variables macroéconomiques à des chocs structurels identifiés par le modèle VAR constituent une référence que tout modèle théorique doit reproduire au mieux. Rotemberg et Woodford (1997) et Christiano et al. (2005) proposent même d’estimer et de tester des modèles DSGE sur la base de ces réponses. Les paramètres du modèle DSGE sont estimés de façon à minimiser un indicateur de distance entre les réponses du modèle théorique à un choc monétaire et celles issues d’un modèle VAR structurel. Dans cette optique, les implications quantitatives des modèles VAR s’avèrent indispensables pour des macroéconomistes qui s’intéressent à la construction et à l’évaluation de modèles dynamiques. De plus, la première génération de modèles DSGE imposait tant de contraintes sur les données qu’ils s’ajustaient mal aux observations. Ceci a longtemps entretenu l’idée qu’ils constituaient essentiellement des curiosités théoriques de peu d’utilité dans une démarche quantitative. Cependant, de nombreux efforts ont été entrepris pour améliorer les spécifications existantes de sorte que ces modèles se sont progressivement imposés comme des outils utiles pour la prévision et l’analyse quantitative de la politique économique. Ces modèles ont ainsi gagné en crédibilité auprès de nombreuses institutions, notamment auprès des banques centrales. Smets et Wouters (2003) ont ainsi développé un modèle DSGE monétaire avec rigidités réelles et nominales dans la lignée de celui proposé par Christiano et al. (2005) et l’ont estimé pour la zone Euro. Un résultat empirique assez remarquable ressort de leur étude : *le modèle DSGE “domine” en termes de prévision un modèle VAR correspondant*. Ce résultat tient en partie au fait que le modèle proposé par Smets et Wouters est d’une grande taille et fait intervenir de très nombreux paramètres.² Plus encore, le nombre de chocs exogènes de ce modèle excède largement le nombre de variables observables. L’expérience montre alors que la densité marginale du modèle diminue avec le nombre de chocs (voir Del Negro, Schorfheide, Smets et Wouters (2006)). Cependant, l’idée que les modèles VAR continuent de dominer des modèles DSGE de taille modeste incluant un petit nombre de frictions continue à persister.

L’objet de cet article est précisément de réévaluer cette croyance et de montrer que des modèles DSGE réels incorporant un seul mécanisme intertemporel supplémentaire (et donc un seul paramètre supplémentaire) sont capables de fournir un ajustement aux données similaire à celui d’un modèle VAR. Il ne s’agit pas ici de “ressusciter” le modèle cano-

¹Un premier exemple connu et bien admis est l’identification des réponses de l’économie à des chocs sur la politique monétaire. Les résultats empiriques issus des modèles VAR ont fortement influencé le développement de modèles théoriques avec des rigidités nominales. Un autre exemple plus récent et controversé est l’identification d’une amélioration technologique sur l’emploi.

²La qualification de la taille d’un modèle reste forcément arbitraire. Par grande taille, nous entendons relativement au modèle DSGE de référence qui ne fait rarement intervenir plus de dix paramètres. A titre de comparaison, le modèle de Smets et Wouters (2003) fait intervenir plus de trente de paramètres.

nique des cycles réels (King et Rebelo (2000)), mais seulement de déterminer le degré de rigidité réelle supplémentaire permettant à ce type de modèle de fournir un ajustement satisfaisant aux données. Nous considérons différents modèles dynamiques concurrents : un modèle VAR, un modèle DSGE sans frictions, un modèle DSGE avec coûts d’ajustement sur l’investissement et un modèle DSGE avec persistance des habitudes de loisir. Ces modèles sont estimés par une méthode de maximum de vraisemblance sur données américaines incluant le taux de croissance de la productivité horaire et les heures travaillées sur la période 1959 :I–2003 :IV. Il ressort de nos estimations que le modèle sans frictions est rejeté au profit d’un modèle intégrant des coûts à l’ajustement ou un mécanisme de persistance des habitudes. De plus, les deux modèles avec frictions fournissent une log-vraisemblance très proche de celle du modèle VAR. Différentes statistiques et critères de choix de modèles concurrents viennent confirmer ces résultats. Il apparaît même que ces modèles DSGE dominent dans certains cas le modèle VAR, notamment pour la prévision hors échantillon. Ces résultats empiriques suggèrent que les modèles DSGE doivent être préférés aux modèles VAR, notamment lorsqu’il s’agit d’identifier les chocs structurels et leurs effets sur la dynamique agrégée.

L’article est organisé comme suit. Dans une première section, nous exposons brièvement les modèles dynamiques considérés dans cette étude. La seconde section présente la méthode d’estimation et les différents critères statistiques de choix entre modèles concurrents. Les résultats empiriques sont exposés dans une troisième section. La quatrième section discute l’identification des chocs structurels. Une dernière section conclue.

2 Les Modèles Concurrents

Cette section présente les différents modèles utilisés pour l’analyse de la dynamique jointe de la productivité et des heures travaillées. Nous retenons essentiellement deux approches. La première, l’analyse VAR, s’inscrit dans la lignée des travaux économétriques initiées par Sims (1980) et se veut descriptive et agnostique. La seconde se revendique structurelle et théoriquement fondée.

2.1 Le modèle VAR

La modélisation VAR s’attache essentiellement à révéler la structure d’autocorrélation des données à des fins essentiellement prédictives. Il s’agit donc de caractériser au mieux la dynamique jointe des variables modélisées — ici le taux de croissance de la productivité et les heures travaillées — afin de révéler la loi les ayant engendrées. La démarche est exclu-

sivement descriptive et a-théorique. L'objectif est donc de maximiser le pouvoir prédictif du modèle.

Soit le processus centré $\{Y_t\}_{t=1}^T$, où Y_t est un vecteur de dimension n_y , la représentation VAR de Y_t est simplement fournie par

$$A(L)Y_t = \varepsilon_t,$$

où $A(L) = (I - A_1L - \dots - A_\ell L^\ell)$ est un polynôme en l'opérateur retard L , et ℓ désigne le nombre de retard du VAR. Les racines du polynôme matriciel $A(L)$ sont toutes supposées de module supérieur à l'unité. ε_t désigne le vecteur des innovations canoniques du processus. ε_t est de moyenne nulle et de matrice de variance covariance $E(\varepsilon_t \varepsilon_t') = \Sigma$.

2.2 Un modèle sans frictions

Nous présentons ici une version minimale d'un modèle d'équilibre général stochastique sous anticipations rationnelles. Nous considérons une économie fermée peuplée par une infinité d'agents rationnels à durée de vie infinie.³ Les agents sont tous supposés identiques et font face à la même structure d'information et aux mêmes sources de chocs structurels. Nous faisons donc l'hypothèse d'existence d'un agent représentatif dont les préférences intertemporelles sont représentées par la fonction d'utilité

$$E_t \sum_{i=0}^{\infty} \beta^i \{ \log(C_{t+i}) - \exp(-\chi_{t+i}) N_{t+i} \}$$

où $\beta \in (0, 1)$ est le taux d'actualisation subjectif, C_t la consommation réelle et N_t les heures travaillées. E_t représente l'opérateur d'espérance conditionnelle à l'ensemble d'information de la période t , c'est à dire au moment où les ménages doivent prendre leur décision de consommation et d'offre de travail. Suivant Hansen (1985), la décision d'offre de travail ne porte pas sur un nombre d'heures variable, mais sur la probabilité variable de travailler un nombre d'heures donné. Le travail est donc indivisible. La situation de chaque agent est alors déterminée par une loterie avec assurance complète du revenu, de sorte que la désutilité du travail est linéaire (Voir Hairault (1992)). L'utilité des agents est soumise à des chocs stochastiques, χ_t , déplaçant l'offre de travail du ménage. Ce choc sur les préférences est supposé suivre un processus stationnaire autorégressif d'ordre un

$$\chi_t = \rho \chi_{t-1} + \sigma_\chi \varepsilon_{\chi,t}$$

avec $|\rho| < 1$, $\sigma_\chi > 0$ et $\varepsilon_{\chi,t}$ est un processus iid de moyenne nulle et de variance unitaire.

³La population est supposée constante et normalisée à l'unité.

De la même façon que nous avons fait l'hypothèse d'existence d'un ménage représentatif, et pour les mêmes raisons, nous supposons qu'il existe une entreprise représentative produisant un bien final homogène en quantité Y_t . La production de ce bien est assurée par une technologie à rendements constants mettant en œuvre le capital physique, K_t , accumulé dans l'économie et les heures offertes par le ménage, N_t . La technologie est représentée par la fonction de production Cobb–Douglas

$$Y_t = K_t^\alpha (\exp(z_t)N_t)^{1-\alpha}$$

où $\alpha \in (0, 1)$. La variable z_t représente un choc sur la productivité globale des facteurs. Ce choc est supposé suivre un processus de type marche aléatoire avec dérive de la forme

$$z_t = z_{t-1} + \gamma_z + \sigma_z \varepsilon_{z,t}$$

avec $\sigma_z > 0$, et $\varepsilon_{z,t}$ est un processus iid de moyenne nulle et de variance unitaire. La variable $\varepsilon_{z,t}$ est supposée orthogonale à $\varepsilon_{\chi,t-q}$ quel que soit q . Le terme constant γ_z représente la dérive dans le processus marche aléatoire et capture la composante déterministe de la croissance. Le bien homogène Y_t peut être utilisé à des fins de consommation (C_t) et d'investissement (I_t). Ce dernier donne lieu à l'accumulation du capital physique utilisé dans le processus de production. Cette accumulation est représentée, de façon standard, par une loi d'inventaire permanent de la forme

$$K_{t+1} = I_t + (1 - \delta)K_t$$

où $\delta \in (0, 1)$ est le taux de dépréciation constant du capital.

La condition d'équilibre sur le marché des biens s'écrit :

$$Y_t = C_t + I_t + G_t$$

où G_t représente les dépenses publique financée par une taxe forfaitaire. Sans perte de généralité, la composante stationnaire des dépenses publiques est supposée constante au cours du cycle, *i.e.* $G_t / \exp(z_t) = \bar{G}$.⁴

Ce modèle servira de référence dans la suite de l'analyse. Il intègre en effet les éléments constitutifs de la vaste majorité des modèles d'équilibre général dynamique utilisés à des fins de prévision : caractère intertemporel des décisions, rationalité des comportements, accumulation du capital, chocs stochastiques affectant l'environnement économique des agents. Il importe toutefois de noter que les propriétés dynamiques de ce modèle sont relativement pauvres. Il est en particulier établi (cf. Cogley et Nason (1995)) que ce modèle

⁴L'effet d'une part variable et stochastique stationnaire des dépenses publiques est similaire, dans ce modèle, à celui d'un choc sur les préférences.

possède un mécanisme de propagation des chocs très faible et engendre peu de persistance des fluctuations. Aussi, diverses voies ont été empruntées par la littérature pour enrichir la dynamique réelle du modèle. Nous en présentons deux qui ont été largement retenues en macroéconométrie appliquée.

2.3 Extensions

Nous considérons deux voies d'enrichissement des mécanismes de propagation des chocs. La première a récemment été introduite par Christiano et Fisher (2003), Christiano et al. (2005) et Eichenbaum et Fisher (2005). Elle repose sur l'idée que la dynamique de cette classe de modèles est largement déterminée par l'accumulation du stock de capital physique. L'accumulation de capital est régie par la relation

$$K_{t+1} = \left[1 - \mathcal{S} \left(\frac{I_t}{I_{t-1}} \right) \right] I_t + (1 - \delta)K_t$$

où la fonction $\mathcal{S}(\cdot)$ prend en compte l'existence de coûts à l'ajustement. $\mathcal{S}(\cdot)$ est une fonction croissante strictement convexe. Elle satisfait aux conditions suivantes : (i) $\mathcal{S}(\gamma_z) = \mathcal{S}'(\gamma_z) = 0$ et (ii) $\varphi = \mathcal{S}''(\gamma_z)\gamma_z^2 > 0$. En conséquence, l'état stationnaire du modèle ne dépend pas de ce paramètre. En revanche, la dynamique du modèle est étroitement liée aux valeurs prises par φ . Il importe de noter que c'est parce que les coûts à l'ajustement portent sur les flux de capital plutôt que sur le stock de capital, que la persistance du modèle est affectée. Dans le cas où les coûts portent sur le stock de capital, seule la volatilité de l'investissement est modifiée et la persistance n'est que très peu altérée. Ceci se comprend aisément dans la mesure où la spécification retenue ici pénalise les variations du niveau d'investissement plutôt que celles du stock de capital.

La seconde extension que nous considérons porte sur un autre comportement clé du modèle : l'offre de travail. Cette modification déjà présente dans l'article originel de Kydland et Prescott (1982)⁵ consiste en l'introduction d'un phénomène de persistance des habitudes dans la consommation de loisir. Plusieurs modélisations de ce phénomène peuvent être envisagées. Nous retenons une version simplifiée de Bouakez et Kano (2006).⁶ Les préférences du ménages sont alors représentées par

$$E_t \sum_{i=0}^{\infty} \beta^i \left\{ \log(C_{t+i}) - \frac{\exp(-\chi_{t+i})}{1 + \nu} \left(\frac{N_{t+i}}{N_{t+i-1}^\psi} \right)^{1+\nu} \right\}$$

⁵Dans leur article, Kydland et Prescott (1982) calibre le processus de persistance des habitudes de façon à renforcer la substitution intertemporelle dans l'offre de travail. Cette approche a été remise en cause notamment par Eichenbaum, Hansen et Singleton (1988) qui établissent, sur données agrégées américaines, que l'offre de travail est sujette à une forte complémentarité intertemporelle.

⁶Ces auteurs considèrent en effet une modélisation avec formation d'un stock d'habitudes.

où $\psi \in (0, 1)$ est le paramètre réglant le degré d'habitudes. Il apparaît ainsi que les décisions d'offre de travail ne sont plus séparables temporellement. Une conséquence directe de cette hypothèse est que la décision d'offre de travail dépend des décisions passées, présentes et futures anticipées comme le montre l'équation définissant le comportement optimal

$$\exp(-\chi_t)N_t^\nu N_{t-1}^{-\psi(1+\nu)} - \beta\psi E_t \left[\exp(-\chi_{t+1})N_{t+1}^{1+\nu} N_t^{-1-\psi(1+\nu)} \right] = \frac{W_t}{C_t}$$

où W_t désigne le taux de salaire réel. L'existence d'un phénomène d'habitude portant sur le loisir crée donc de la persistance dans l'offre de travail dans la mesure où l'agent cherche à lisser son comportement en termes de loisir, et ainsi le profil temporel du temps qu'il consacre aux activités productives. Cette persistance se transmet alors directement à l'activité de production, et est partiellement renforcée par l'accumulation de capital.

Dans un premier temps, chaque modèle est stationnarisé par sa composante de croissance $\exp(z_t)$. Leur résolution est ensuite assurée en log-linéarisant le système d'équations aux différences finies qui en caractérise la dynamique autour de son état stationnaire. Le système linéarisé est alors résolu par une technique standard et fournit une représentation espace-état du modèle de la forme

$$Y_t = M_Y X_t \tag{1}$$

$$X_{t+1} = M_X X_t + M_E \varepsilon_{t+1} \tag{2}$$

où $\varepsilon \rightsquigarrow \mathcal{N}(0, \Sigma)$. Y_t désigne le vecteur de dimension $(n_y \times 1)$ constitué par les variables directement observables par l'économètre — dans notre cas les heures travaillées et le taux de croissance de la productivité. X_t est le vecteur $(n_x \times 1)$ des variables d'état du modèle : $X_t \equiv (K_t, z_t, \chi_t)'$ pour le modèle sans friction, et $X_t \equiv (K_t, \Upsilon_t, z_t, \chi_t)'$ pour le modèle avec friction ($\Upsilon_t = I_{t-1}$ dans le modèle avec coûts d'ajustement, et $\Upsilon_t = N_{t-1}$ pour le modèle avec persistance des habitudes). Enfin, ε_t est le vecteur $(n_\varepsilon \times 1)$ des innovations des chocs $(\varepsilon_{z,t}, \varepsilon_{\chi,t})$.

3 Estimation et Critères de Choix de Modèles

Nous présentons la méthode d'estimation et de tests des modèles dynamiques concurrents de la section précédente. La méthode d'estimation retenue est, pour chaque modèle, le maximum de vraisemblance à information complète. A partir de la valeur estimée des paramètres, nous considérons ensuite différents critères de choix de modèles.

3.1 Méthode d'estimation

Nous exposons ici brièvement la mise en œuvre de la méthode d'estimation par maximum de vraisemblance des modèles de cycle proposés dans la section précédente.⁷ La solution obtenue par approximation log-linéaire des différents modèles d'équilibre général dynamiques s'exprime sous la forme générique suivante

$$Y_t = M_Y(\theta)X_t \quad (3)$$

$$X_{t+1} = M_X(\theta)X_t + M_E(\theta)\varepsilon_{t+1} \quad (4)$$

où $\varepsilon \rightsquigarrow \mathcal{N}(0, \Sigma)$. θ désigne le vecteur, de taille $(n_\theta \times 1)$, des paramètres à estimer. Dans ce qui va suivre, seul un sous-ensemble des paramètres des modèles est estimé. En conséquence, le vecteur θ est composé de deux sous-vecteurs de paramètres $\theta = (\theta_1, \theta_2)$, où θ_1 est étalonné avant l'estimation et seul θ_2 est réellement estimé. En conséquence, le vecteur de paramètres θ_2 que l'on cherche à estimer est de dimension $n_{\theta_2} < n_\theta$. Pour simplifier la présentation, nous n'effectuons pas cette distinction par la suite. $M_Y(\theta)$, $M_X(\theta)$, $M_E(\theta)$ sont des matrices de taille conformes aux dimensions des vecteurs dont les coefficients sont des fonctions non-linéaires compliquées des paramètres à estimer.

L'économètre observe le vecteur de données $\{Y_t\}_{t=1}^T$, que nous supposons ici centré et stationnaire. A partir de la représentation espace-état de la solution de chaque modèle, et sachant que les innovations ε_t sont normalement distribuées, la distribution conditionnelle du vecteur de variables observables Y_t est normale

$$Y_t | \mathcal{Y}_{t-1} \rightsquigarrow \mathcal{N}(M_Y(\theta)X_{t|t-1}, M_Y(\theta)P_{t|t-1}M_Y(\theta)')$$

où $\mathcal{Y}_{t-1} = \{Y_{t-i}; i = 1, \dots\}$. $X_{t|t-1} = E(X_t | \mathcal{I}_{t-1})$ et $P_{t|t-1} = E((X_t - X_{t|t-1})(X_t - X_{t|t-1})')$, où \mathcal{I}_{t-1} représente l'ensemble d'information utilisé pour la prévision. Nous pouvons alors écrire la densité conditionnelle de Y_t pour une date quelconque $t \in (1, T)$

$$f(Y_t | \mathcal{Y}_{t-1}; \theta) = (2\pi)^{-n_y/2} |M_Y(\theta)P_{t|t-1}M_Y(\theta)'|^{-1/2} \\ \times \exp \left[-\frac{1}{2} (Y_t - M_Y(\theta)X_{t|t-1})' (M_Y(\theta)P_{t|t-1}M_Y(\theta)')^{-1} (Y_t - M_Y(\theta)X_{t|t-1}) \right]$$

et en déduire la log-vraisemblance de l'échantillon de données $t = 1, \dots, T$

$$\log(\mathcal{L}(\theta | \mathcal{Y})) = \sum_{t=1}^T \log(f(Y_t | \mathcal{Y}_{t-1}; \theta))$$

Pour une valeur donnée des paramètres θ des modèles, l'évaluation de la log-vraisemblance nécessite la connaissance de deux quantités $X_{t|t-1}$ et $P_{t|t-1}$. $X_{t|t-1}$ correspond à la prévision

⁷Il faut noter que cette méthode s'applique également au modèle VAR.

à un pas de l'état du système, tandis que $P_{t|t-1}$ est la matrice de variance covariance de l'erreur de prévision associée. Chacune de ces deux quantités peut être obtenue par application du filtre de Kalman.⁸ Les paramètres des différents modèles sont alors estimés en maximisant la log-vraisemblance à l'aide d'algorithmes usuels d'optimisation. Lorsque l'algorithme a convergé, les écart-types des paramètres estimés peuvent être obtenus en calculant la covariance du gradient.

Concernant le modèle VAR, les paramètres peuvent être plus facilement obtenus à l'aide des moindres carrés ordinaires appliqués à chaque équation. En effet, en l'absence de restrictions inter-équations sur les matrices du système autorégressif, A_i pour $i = 1, \dots, \ell$, et la matrice de variance-covariance des erreurs, Σ , l'estimation par les moindres carrés ordinaires équation par équation est équivalente à l'estimation par le maximum de vraisemblance sur l'ensemble du système (Gouriéroux et Monfort (1990)).⁹

3.2 Critères de Choix de Modèles et Tests

Nous présentons ici différents critères de choix de modèles concurrents non-emboîtés. Ces différents critères sont des fonctions réelles définies pour les modèles dynamiques en compétition sur l'échantillon ayant servi à l'estimation $\{Y_t\}_{t=1}^{i=T}$. Le choix peut alors s'effectuer en regardant, pour un ou tous les éléments de $\{Y_t\}_{t=1}^{i=T}$, le modèle qui maximise ou minimise un critère donné. Ces critères sont évalués à convergence, $\hat{\theta}_T$, de la procédure d'estimation. Ces différents critères ont des performances convenables dans la pratique, même si leurs fondements théoriques ne sont pas complètement établis (Gouriéroux et Monfort (1989))

A partir de l'estimation, il est possible de comparer les log-vraisemblances des différents modèles en compétition. Cette comparaison permet alors de sélectionner un modèle sur sa capacité à s'ajuster aux données. Cependant, ce critère n'intègre pas le caractère parcimonieux du modèle. Avec le critère d'information d'Akaike (*AIC*), le maximum de la log-vraisemblance est pénalisé d'une valeur égale au nombre de paramètres estimés. Le modèle choisi est alors celui qui donne le maximum de

$$AIC = \log(\mathcal{L}(\hat{\theta}_T|\mathcal{Y})) - n_\theta$$

Différentes modifications du critère d'Akaike ont été proposées. Selon le critère de Schwarz, le modèle choisi est celui donnant le maximum de

$$BIC = \log(\mathcal{L}(\hat{\theta}_T|\mathcal{Y})) - \frac{n_\theta}{2} \log T$$

⁸Voir l'annexe A pour plus de détails.

⁹Tout modèle VAR d'ordre ℓ peut se réécrire sous la forme (3)–(4) et alors être estimé en maximisant la log-vraisemblance.

Il apparaît ici que le critère *BIC* pénalise d’avantage le nombre de paramètres que le critère *AIC* lorsque la taille de l’échantillon est suffisamment grande ($T > 8$), ce qui est vérifié en pratique. De la même façon, le critère de Hannan et Quinn pénalise d’avantage encore les modèles non parcimonieux

$$HQ = \log(\mathcal{L}(\hat{\theta}_T|\mathcal{Y})) - n_\theta \log \log T$$

Lorsque le modèle est linéaire dans les variables, ce qui est le cas pour tous les modèles envisagés dans l’article, un critère classique simple est le coefficient de détermination. En notant $\hat{Y}_{i,t} = M_{I,Y}(\hat{\theta}_T)X_{t|t-1}$ le vecteur de valeurs ajustées pour la variable $i = 1, \dots, n_y$ sélectionnée dans Y_t , le coefficient de détermination est donnée par $R_i^2 = \sum_{t=1}^T \hat{Y}_{i,t}^2 / \sum_{t=1}^T Y_{i,t}^2$. Plus la prévision de la variable $Y_{i,t}$ est “proche” de sa réalisation, plus le coefficient de détermination est proche de l’unité. On pourra alors retenir le modèle ayant le meilleur coefficient de détermination, c’est à dire celui fournissant les meilleures prévisions $Y_{i,t}$. Un avantage de cette approche est de permettre de juger la qualité d’ajustement du modèle aux données pour les variables observables. Elle permet ainsi de préciser les éventuels échecs de la modélisation proposée. Dans le cas où les erreurs sont normalement distribuées, on peut utiliser le coefficient de détermination pour construire une statistique de Fisher qui correspond à l’absence de pouvoir explicatif du modèle concernant la variable $Y_{i,t}$. Cette statistique est donnée par $F_i = (T - \dim \theta)R_i^2 / (\dim \theta(1 - R_i^2))$. Elle est une fonction croissante de R_i^2 et elle permet de tester simplement le pouvoir prédictif de différents modèles. Le coefficient de détermination simple augmente avec le nombre de variables explicatives, ce qui limite son utilisation comme critère de choix entre différents modèles linéaires. En pratique, on peut lui préférer un coefficient ajusté par le nombre de paramètres. Celui-ci est défini par $\bar{R}_i^2 = 1 - (1 - R_i^2) (T/(T - n_\theta))$.¹⁰

Nous complétons finalement ces différents critères en évaluant les propriétés du terme d’erreur de prévision à un pas d’une variable i de $Y_{i,t}$ ($\hat{\varepsilon}_{i,t} = Y_{i,t} - \hat{Y}_{i,t}$). Plus précisément, il s’agit d’évaluer l’absence d’autocorrélation dans le processus de ce terme d’erreur estimé. En effet, la présence d’erreurs autocorrélées est révélatrice d’une erreur de spécification tant dans les modèles de cycle — mécanismes internes de persistance insuffisants — que dans le modèle VAR — nombre insuffisant de retards. Pour tester l’indépendance temporelle des erreurs, nous utilisons la statistique de Box–Pierce, BP(k), et de Lyung–Box, LB(k), où k est le nombre de retards retenu. Sous l’hypothèse nulle d’absence d’autocorrélation

¹⁰L’interprétation du R^2 (ou du \bar{R}^2) est toutefois délicate. En effet, seule une approximation (log-)linéaire du modèle structurel est estimée. Sous l’hypothèse que l’approximation soit de bonne qualité, l’utilisation du R^2 est justifiée. Toutefois, il convient de garder à l’esprit que le vrai processus générateur des données est un modèle dynamique non-linéaire dans les variables. Dans ce contexte, le coefficient de détermination doit être interprété avec prudence. Cette même remarque s’applique pour l’utilisation de la statistique de Fisher.

ces statistiques sont distribués selon une loi de Chi-deux à k degrés de liberté.

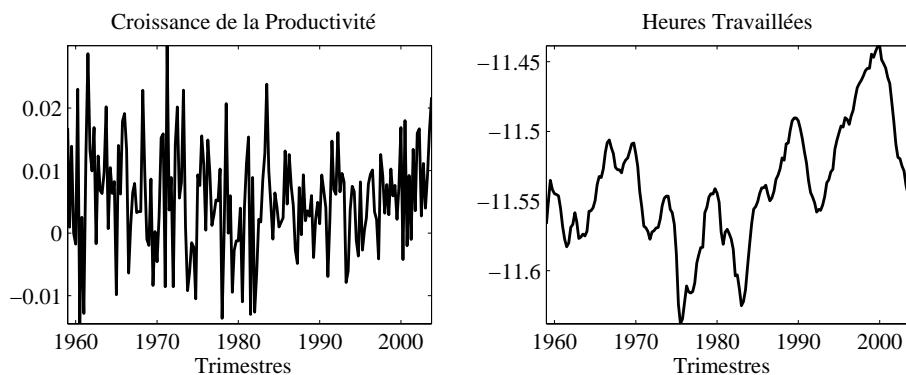
4 Résultats

Cette section présente les résultats d'estimation des modèles proposés et une évaluation de leurs qualités prédictives.

4.1 Estimation

L'étude est menée sur données trimestrielles américaines couvrant la période 1959 :I–2002-IV (voir figure 1). Les modèles d'équilibre général dynamique rencontrent peu de difficulté à rendre compte des fluctuations agrégées sur le marché des biens. En revanche, leur échec à expliquer les fluctuations sur le marché du travail est maintenant bien connu et documenté. En particulier, ils échouent dans la reproduction des comouvements entre heures travaillées et productivité du travail. Il est alors naturel d'évaluer les différents modèles sur leur capacité à rendre compte de cette dynamique. Aussi, dans la lignée des travaux de Galí et Rabanal (2004), la croissance de la productivité est mesurée par le taux de croissance de la productivité horaire dans l'industrie et les services. Les heures travaillées sont ici définies par le produit du taux d'emploi¹¹ et des heures effectives individuelles. Cette définition est conforme à celle du modèle (voir Christiano, Eichenbaum et Vigfusson (2004)). Le produit par tête peut alors simplement être obtenu en multipliant la productivité horaire par les heures ainsi définies.

FIG. 1 – Données



La présence des heures travaillées pose un problème spécifique soulevé par plusieurs études : celui de la non-stationnarité des heures. Nous suivons la recommandation de Christiano

¹¹Le taux d'emploi est ici mesuré par l'emploi rapporté à la population civile non-institutionnelle âgée de 16 ans et plus.

et al. (2004) et utilisons les heures en niveau lors de l'estimation du VAR. Deux raisons motivent ce choix. La première est que les résultats ne sont que très partiellement affectée par ce changement de représentation en pratique. La seconde est plus structurelle et renvoie au modèles théoriques proposés. Dans chacun de ces modèles, les heures travaillées sont, tout comme dans les données, mesurées par le nombre d'heures travaillées par tête. Il s'agit ainsi d'une quantité bornée qui ne peut pas, asymptotiquement, présenter de dérive. Les heures sont donc, au moins dans le modèle théorique, stationnaires. Les données sont centrées. L'application de tests du rapport de vraisemblance nous conduit, sans ambiguïté, à retenir 4 retards dans le VAR.

L'estimation des modèles structurels est réalisée pour la même période. Nous n'estimons pas l'ensemble des paramètres des modèles. En effet, les modèles DSGE sont sujets à des problèmes d'estimations largement liés à des problèmes d'identification de l'ensemble des paramètres. Aussi, afin de minimiser l'occurrence de tels problèmes nous avons pris le parti de n'estimer que les paramètres liés aux processus exogènes, et ceux ne pouvant pas être déterminés en utilisant les restrictions imposées par l'état stationnaire du modèle. Ceci nous a conduit à séparer l'ensemble des paramètres du modèle en deux parties $\theta_1 = \{\alpha, \beta, \delta, \gamma_z, g/y, \nu\}$ et $\theta_2 = \{\rho_\chi, \sigma_\chi, \sigma_z, (\varphi \text{ ou } \psi)\}$. La première, collectée dans le vecteur θ_1 , est fixée en référence à des études antérieures ou de façon à reproduire certains ratios de long terme observés. (Voir tableau 1)¹² Ainsi, α est fixé de façon à reproduire la part

TAB. 1 – Paramètres fixés

γ_z	α	δ	β	ν	g/y
1.005	0.360	0.025	0.990	0/1	0.200

de la rémunération salariale dans la valeur ajoutée américaine sur la période considérée (64%). Le facteur d'escompte, β , est fixé à 0.99. Le taux de dépréciation δ est de 2.5% par trimestres, soit un déclassement du capital de 10% par an. γ_z est fixé de façon à reproduire le taux de croissance moyen de la productivité sur la période. De façon similaire, la part des dépenses gouvernementales est étalonnée à sa contrepartie empirique sur la période. Ces paramètres sont communs à toutes les versions du modèle. Enfin, ν est fixé à zéro dans le modèle sans frictions réelles et le modèle à coûts d'ajustement. ν est fixé à l'unité dans le modèle avec persistance des habitudes en référence à l'étude de Bouakez et Kano (2006).

Le tableau 2 reporte les valeurs estimées des paramètres structurels pour les différents

¹²Il est connu depuis les travaux de Cogley et Nason (1995) que la manipulation de ces paramètres ne permet pas d'enrichir les mécanismes de propagation des chocs de ces modèles. L'information contenue dans les fluctuations des autres variables du modèle (consommation, investissement) ne permet pas de lever les problèmes d'identification des paramètres θ_1 .

modèles concurrents, ainsi que la log-vraisemblance associée à chacun des modèles. Une

TAB. 2 – Résultats d’estimation

	σ_γ	ρ_χ	σ_χ	φ	ψ	Log-likelihood
Absence de frictions	0.0142 (0.0007)	0.9874 (0.0092)	0.0070 (0.0008)	–	–	1160.04
Coût à l’ajustement	0.0142 (0.0006)	0.9777 (0.0110)	0.0063 (0.0004)	0.6468 (0.1797)	–	1250.72
Persistance des habitudes	0.0143 (0.0006)	0.9664 (0.0169)	0.0110 (0.0009)	–	0.4872 (0.0368)	1250.76
VAR	–	–	–	–	–	1268.25

Note : Les écarts-types sont reportées entre crochets.

première observation se dégage du tableau 2. L’ensemble des paramètres est estimé avec précision. Ce résultat a deux raisons essentielles. La première tient au faible nombre de paramètres estimés. En restreignant notre attention aux processus des chocs et aux mécanismes de propagation centraux des modèles, nous autorisons une meilleure identification des paramètres. La seconde tient au caractère des mécanismes estimés. Les mécanismes introduits dans les modèles enrichissent la propagation des chocs et la persistance de leurs effets. Or, il est bien connu depuis les travaux de Cogley et Nason (1995) que le modèle de cycle réel canonique présente des mécanismes de propagations trop faibles au regard des données. Il n’est donc pas surprenant que les paramètres les caractérisant apparaissent significatifs lors de leur estimation. Une seconde observation tient à la constance de la volatilité du choc de productivité entre les expériences menées. Quel que soit le modèle considéré, la volatilité du choc s’établit à 1.4% par trimestre. Cette valeur est très proche de la valeur obtenue dans différentes études précédentes (voir Hansen (1997), Chari, Kehoe et McGrattan (2005)). En revanche, la volatilité et la persistance du choc de préférences varie sensiblement entre les modèles. En effet, en l’absence de frictions réelles, la difficulté à reproduire la persistance observée des heures travaillées conduit l’algorithme à augmenter la persistance du choc de préférences, tandis que, pour reproduire leur volatilité, l’écart-type du choc de préférences est sensiblement réduit. Lorsque les mécanismes de propagation sont renforcés dans le modèle (modèle à coûts d’ajustement), la persistance et la volatilité du choc de préférences sont sensiblement réduites. En effet, les coûts d’ajustement lissent les fluctuations de l’investissement et ainsi de la production. La demande de travail est plus lisse et ainsi les heures travaillées à l’équilibre héritent de cette persistance. L’effet est encore renforcé lorsque le mécanisme de persistance porte directement sur les décisions d’emploi (persistance des habitudes). Cependant, dans ce cas, la décision d’offre de travail étant directement lissée, une plus forte volatilité du choc de préférences est nécessaire pour reproduire la volatilité des heures.

Le niveau de la log-vraisemblance atteint par chacun des modèles indique que le modèle VAR semble dominer l'ensemble des modélisations structurelles proposées. Il importe cependant de noter que le modèle VAR nécessite l'estimation de 19 paramètres alors que le plus complet des modèles structurels ne nécessite d'estimer que 4 paramètres. Il convient donc de nuancer quelque peu le succès du VAR, et nous reviendrons sur ce point dans la prochaine section. En revanche, les résultats indiquent clairement que le modèle sans frictions réelles est dominé par les deux autres modèles. La conclusion du test est claire. Le modèle sans frictions est rejeté par les données.¹³ Il est également intéressant de noter que les modèles avec frictions conduisent à une vraisemblance identique. Ceci suggère qu'un modèle introduisant les deux frictions ferait face à de sévères problèmes d'identification, ce qui doit nous conduire à nous interroger sur l'identification des macro-modèles récents introduisant plusieurs rigidités réelles.¹⁴

4.2 Evaluation des modèles

La section précédente a fourni une première évaluation des performances des modèles en termes de pouvoir prédictif. Le modèle VAR semble dominer l'ensemble des modèles structurels concurrents. Cette section propose d'affiner l'analyse précédente en intégrant un critère de parcimonie dans la décision et en évaluant directement le caractère prédictif de chacun des modèles.¹⁵

Dans un premier temps, le tableau 3 revient sur le critère de log-vraisemblance évoqué à la section précédente en proposant un critère tenant compte à la fois du niveau de log-vraisemblance atteint par le modèle et d'un élément de parcimonie. Le tableau reporte donc les critères AIC, BIC et HQ décrits dans la section 3.1. Plusieurs conclusions s'imposent. Tout d'abord, ces indicateurs confirment les mauvais résultats obtenus avec un modèle n'intégrant aucune friction réelle. En revanche, ils nous conduisent à réviser sensiblement notre jugement sur le modèle VAR. Ce dernier est pénalisé par son manque relatif de parcimonie. Si le critère AIC indique que le modèle VAR domine encore les modèles structurels intégrant des frictions réelles, les critères BIC et HQ nous indiquent en revanche que le modèle à frictions réelles domine clairement le modèle VAR. En effet, par construction, ces critères pénalisent plus fortement le manque de parcimonie que le critère

¹³Un test de rapport de vraisemblance montre que la restriction $\varphi = 0$ dans le modèle avec coûts d'ajustement est rejeté par les données.

¹⁴L'estimation d'un modèle combinant persistance des habitudes et coûts à l'ajustement sur l'investissement conduit à une estimation très imprécise du paramètre de coûts à l'ajustement, même si un test de rapport de vraisemblance conduirait à ne pas rejeter la significativité du mécanisme à un seuil de 5% (Vraisemblance obtenue : 1252.74, à comparer à 1250.76 dans le Tableau 2).

¹⁵Une approche complémentaire repose sur la comparaison des modèles par les tests de modèles non-emboîtés (par exemple les tests d'englobement).

TAB. 3 – Qualité de prédiction

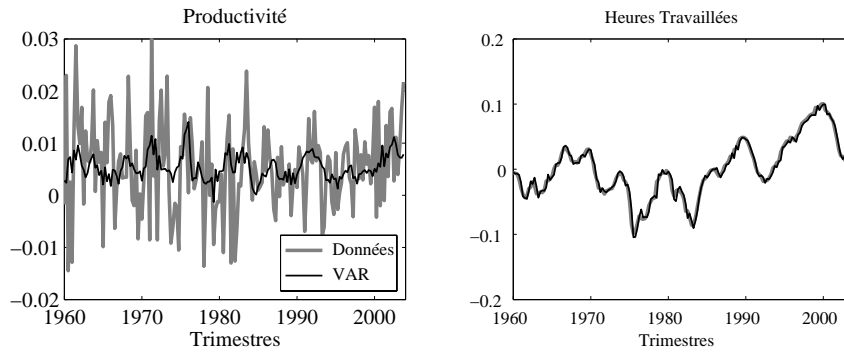
	AIC	BIC	HQ
Absence de frictions	1157.04	1152.29	1155.11
Coût à l’ajustement	1246.72	1240.38	1244.15
Persistance des habitudes	1246.76	1240.42	1244.19
VAR	1249.25	1219.13	1237.04

AIC. Dès lors le modèle VAR, qui nécessite l’estimation de 19 paramètres, est fortement pénalisé relativement au modèle structurel. Les résultats indiquent également qu’il n’est pas possible de discriminer entre les modèles à coûts d’ajustement ou de persistance des habitudes sur la base des critères d’information fondés sur la vraisemblance.

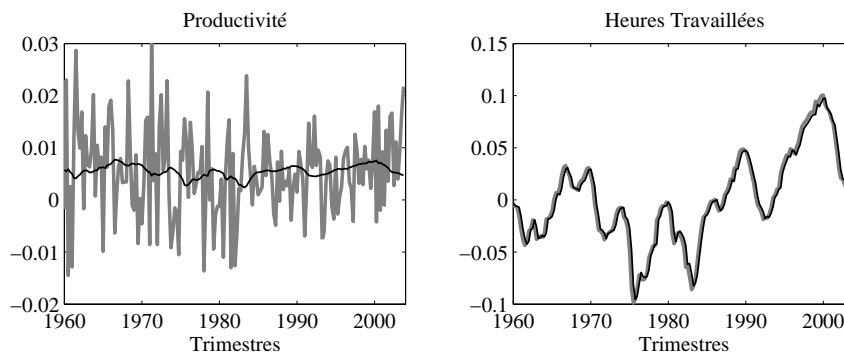
La démarche a-théorique prônée par Sims (1980) en caractérisant la loi jointe des données fournit un outil de prévision puissant. Il est donc particulièrement important d’évaluer le VAR sur ses capacités prédictives. La figure 2 compare les données à leur prévision à un pas réalisée par le VAR ou par le modèle structurel. Les graphiques montrent clairement que l’ensemble des modèles permet de reproduire correctement le profil des heures travaillées. En revanche, le taux de croissance de la productivité prédit à la fois pas le VAR ou les différents modèles structurels apparaît trop lisse au regard des données. Cet excès de lissage est particulièrement flagrant pour le modèle sans frictions réelles. En effet, la modélisation de l’offre de travail à la Hansen (1985) présente la particularité que la productivité du travail est essentiellement proportionnelle à la consommation à l’équilibre. Le modèle s’articulant autour du modèle de revenu permanent, il n’est pas surprenant d’engendrer un profil temporel de la productivité lisse. L’introduction de la persistance des habitudes sur le loisir permet de pallier partiellement cette insuffisance. Le taux de croissance de la productivité présente un profil similaire à celui observé dans le modèle VAR. Il convient toutefois de qualifier ce résultat. Le tableau 4 reporte ainsi des critères prédictif standard — R^2 , \bar{R}^2 , et test de Fisher. Le R^2 des différents modèles montre clairement que tous les modèles permettent de reproduire le profil des heures des façon satisfaisante. L’ensemble des modèles présente un R^2 supérieur à 0.9. Il convient cependant d’insister sur le fait que le modèle sans frictions réelles possède des qualités prédictives inférieures à celles des modèles à friction ou du VAR. Il est également intéressant de noter que le modèle avec persistance des habitudes domine légèrement le modèle VAR, quel que soit le critère retenu. Les qualités prédictives sont en revanche beaucoup plus décevantes lorsque l’attention se porte sur le taux de croissance de la productivité. Aucun modèle ne permet de prédire l’évolution du taux de croissance de la productivité, confirmant ainsi l’impression graphique laissée par la figure 2. Ainsi le modèle VAR présente un R^2 inférieur à 0.1. Le \bar{R}^2 renforce cette conclusion, le VAR étant alors légèrement dominé par l’ensemble

FIG. 2 – Prédiction à 1 pas

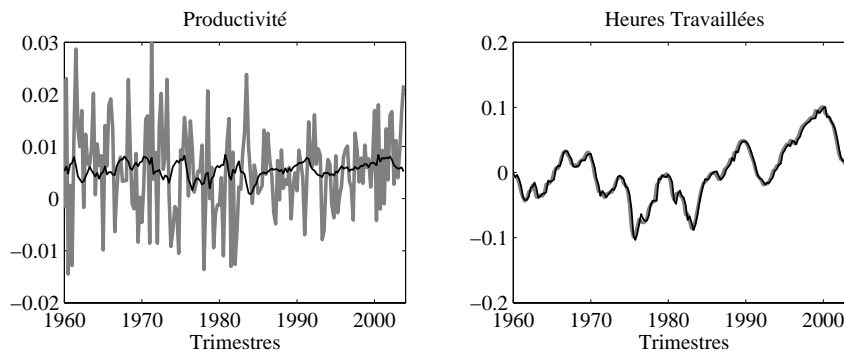
(a) Modèle VAR



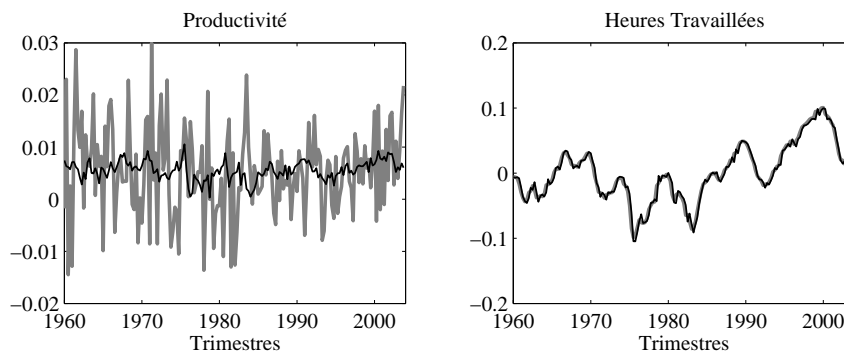
(b) Modèle sans Frictions réelles



(c) Modèle avec coûts d'ajustement



(d) Modèle avec persistance des habitudes



TAB. 4 – Mesure de prédiction

Modèle	Productivité			Heures travaillées		
	R^2	\bar{R}^2	Fisher	R^2	\bar{R}^2	Fisher
Absence de frictions	0.0179	0.0065	1.0486 [0.3725]	0.9363	0.9356	847.5988 [0.0000]
Coût à l’ajustement	0.0319	0.0150	1.4153 [0.2309]	0.9691	0.9685	1347.0780 [0.0000]
Persistance des habitudes	0.0461	0.0295	2.0795 [0.0855]	0.9851	0.9848	2841.6074 [0.0000]
VAR	0.0853	-0.0196	0.7703 [0.7391]	0.9800	0.9777	404.0938 [0.0000]

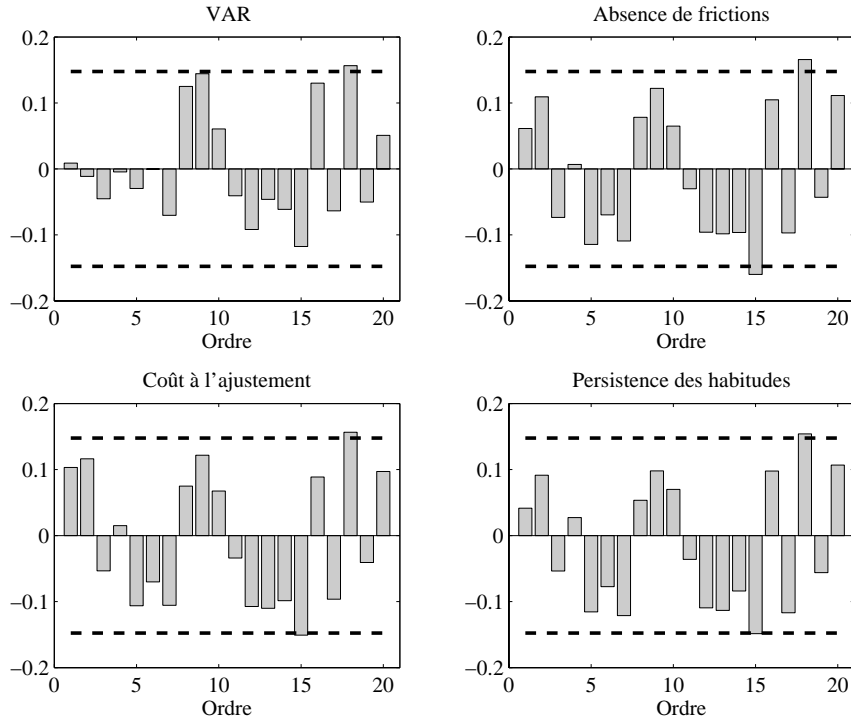
Note : Les p-values sont reportées entre crochets.

des modèles structurels. Le test de Fisher indique alors que seul le modèle avec persistance des habitudes est statistiquement valide au seuil de 10%. Ainsi, les qualités prédictives du modèle structurel apparaissent meilleures que celles délivrées par le modèle VAR, tout en proposant clairement une interprétation structurelle des résultats. Ces résultats suggèrent que la productivité est essentiellement exogènes dans ce modèle. En effet, l’examen de l’équilibre sur le marché du travail nous montre que celle-ci est essentiellement une fonction des chocs, de la consommation et du capital. Le mécanisme de substitution intertemporelle conduit à lisser le comportement de ces deux dernières variables — le lissage du capital étant renforcé par la présence de coûts à l’ajustement. Il n’est donc pas surprenant que la dynamique de la productivité soit essentiellement capturée par l’évolution des chocs, en particulier du choc technologique.

Il importe maintenant de vérifier que les erreurs de prévision ne présentent pas d’auto-corrélation. Une telle évaluation permet en effet de jauger de la bonne spécification — en particulier de la dynamique — des modèle proposés. Les figures 3 et 4 reportent l’autocorrélation des erreurs de prévision à un pas de chacun des modèles pour le taux de croissance de la productivité d’une part et les heures travaillées d’autre part.

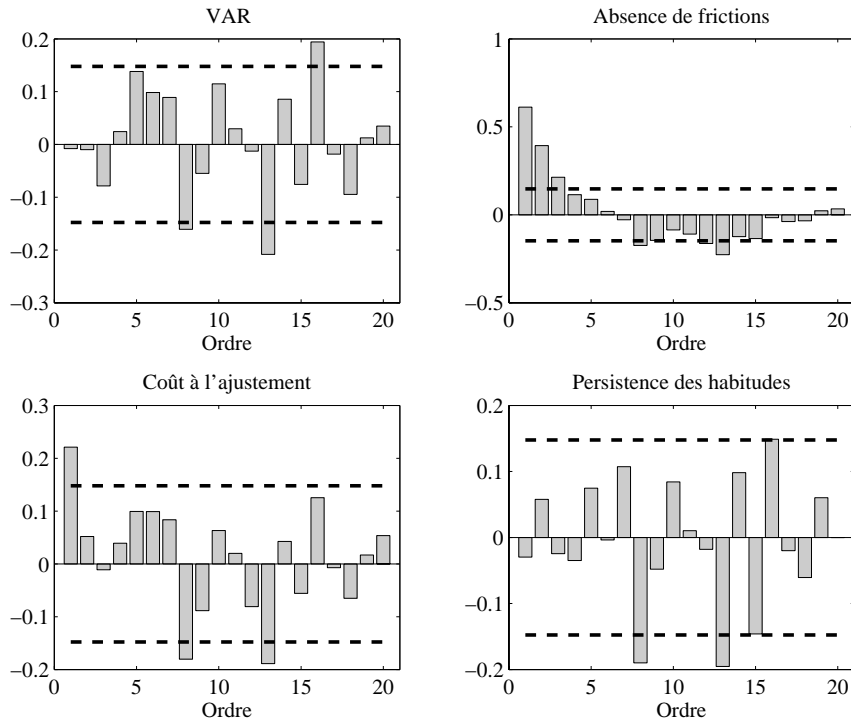
L’examen graphique montre que si les modèles structurels présentent sensiblement plus d’autocorrélation des erreurs que le modèle VAR, cette autocorrélation n’est pas significative au seuil de 5% (seuil en ligne pointillée). Ce résultat est confirmé par le tableau 5 qui reporte les test de Box–Pierce et de Lyung–Box pour différents nombres de retard. Quel que soit le nombre de retards pris en compte pour le calcul de la statistique, les résultats indiquent que tous les modèles présentent des erreurs de prévision du taux de croissance de la productivité non autocorrélées aux seuils de 10% et de 5%. Les seules exceptions sont le modèle structurel sans frictions ou avec coût d’ajustement sur l’investissement lorsque le nombre de retard est important ($k=20$). L’examen du profil d’autocorrélation des erreurs

FIG. 3 – Autocorrélation des erreurs de prévision (Productivité)



Note : Ligne pointillée : Intervalle de confiance à 95%.

FIG. 4 – Autocorrélation des erreurs de prévision (Heures travaillées)



Note : Ligne pointillée : Intervalle de confiance à 95%.

TAB. 5 – Autocorrelation des erreurs de prévision

	Productivité		Heures	
	BP	LB	BP	LB
<i>k</i> = 4				
Absence de frictions	2.9786	3.0462	84.3377	85.9967
	[0.5614]	[0.5501]	[0.0000]	[0.0000]
Coût à l'ajustement	4.1999	4.2865	5.5592	5.6667
	[0.3796]	[0.3686]	[0.2346]	[0.2255]
Persistance des habitudes	1.7726	1.8125	1.3954	1.4222
	[0.7775]	[0.7702]	[0.8450]	[0.8403]
VAR	0.4008	0.4121	1.2207	1.2563
	[0.9824]	[0.9815]	[0.8747]	[0.8687]
<i>k</i> = 12				
Absence de frictions	14.5119	15.2717	99.4736	102.1631
	[0.2692]	[0.2269]	[0.0000]	[0.0000]
Coût à l'ajustement	16.1099	16.9213	17.1083	17.8933
	[0.1863]	[0.1526]	[0.1456]	[0.1190]
Persistance des habitudes	13.6290	14.3803	12.0689	12.7166
	[0.3250]	[0.2771]	[0.4402]	[0.3900]
VAR	10.2618	10.9271	15.2698	16.0860
	[0.5930]	[0.5352]	[0.2270]	[0.1873]
<i>k</i> = 20				
Absence de frictions	31.0073	33.6240	116.0165	120.3102
	[0.0551]	[0.0288]	[0.0000]	[0.0000]
Coût à l'ajustement	30.9903	33.4667	29.4722	31.6003
	[0.0553]	[0.0300]	[0.0789]	[0.0477]
Persistance des habitudes	29.1187	31.6179	27.4969	29.8178
	[0.0854]	[0.0475]	[0.1219]	[0.0729]
VAR	22.6217	24.7320	33.7095	36.4487
	[0.3077]	[0.2119]	[0.0282]	[0.0136]

Note : BP et LB désignent respectivement les statistiques de Box–Pierce et de Lyung–Box pour un nombre de retards égal à *k*. Les p–values sont reportées entre crochets.

de prévision des heures travaillées par le modèle sans friction démontre clairement l'échec du modèle structurel. Le profil de la fonction d'autocorrélation est proche de celui d'un processus à composante AR(1) de coefficient de persistance égal à 0.61. Ainsi, le modèle échoue à prévoir les heures travaillées à un pas. Cet échec n'est pas surprenant dans la mesure où 4 retards sont nécessaires pour blanchir les résidus du VAR, alors que le modèle sans frictions réelles ne possède pas de mécanismes de propagation et de persistance importants. Dès que la dynamique du modèle est enrichie par des mécanismes de persistance, la fonction d'autocorrélation des erreurs de prévision présente un profil semblable à celui obtenu pour le modèle VAR. En particulier, le modèle VAR, tout comme les modèles structurels avec frictions, présente une autocorrélation significative aux ordres 8, 13 et 16. Le modèle à coûts à l'ajustement présente une autocorrélation à l'ordre 1 significativement différente de zéro (0.22), indiquant la présence d'une composante MA(1) dans le processus des erreurs de prévision. Cette composante est corrigée par le modèle avec persistance des habitudes qui présente un profil d'autocorrélation similaire à celui du VAR. Le modèle de persistance des habitudes supplante légèrement le modèle VAR pour des ordres d'autocorrélation élevés. Ainsi, les statistiques de Box–Pierce et de Lyung–Box conduisent au rejet de l'hypothèse nulle d'absence d'autocorrélation des erreurs de prévision du VAR pour un ordre élevé de retards. En revanche, le modèle avec persistance des habitudes ne rejette pas l'hypothèse nulle.

L'exercice de prévision est maintenant étendu à des prévisions hors échantillon. Chaque modèle est estimé de façon récursive vers l'avant. Un exercice de prévision à 1, 4 et 8 trimestres est alors mené pour chaque estimation à partir du point terminal de chacun des sous échantillons. La date terminale varie entre 1997 :I et 1999 :IV. Le tableau 6 reporte alors les moyennes de l'erreur moyenne (EM), de l'erreur absolue moyenne (EAM) et de l'erreur quadratique moyenne (EQM) associées à chaque modèle pour un horizon de $k=1, 4$ et 8 périodes pour les douze échantillons sur lesquels chaque modèle a été estimé. L'évolution de la productivité est toujours mieux prédite par les modèles structurels. En particulier les modèles avec frictions — modèle à coûts d'ajustement (MCI) et modèle avec persistance des habitudes (MHP) — dominant clairement le modèle VAR pour la prévision de la productivité. Ceci est vrai quel que soit l'horizon de prévision et quel que soit le critère d'évaluation retenu. Ce résultat doit être sensiblement mitigé en ce qui concerne les heures travaillées. Le modèle VAR présente en effet des performances supérieures pour la prévision à un trimestre de l'évolution des heures travaillées. En revanche, les modèles DSGE avec frictions permettent une prévision de meilleure qualité à des horizons supérieurs. En effet, les modèles DSGE tire ici avantage de leur cohérence interne en termes de propriété de long terme.

TAB. 6 – Pr evision k pas

		Productivit�e			
Horizon	Erreur	MSF	MCI	MHP	VAR
k=1	EM	-1.0545e-003	-3.9209e-004	-5.2968e-004	2.1569e-003
	EAM	3.4821e-003	3.4417e-003	3.7002e-003	4.0745e-003
	EQM	1.9547e-005	1.7839e-005	1.9116e-005	2.4148e-005
k=4	EM	6.2687e-005	2.4887e-004	-2.4530e-004	2.7822e-003
	EAM	4.0132e-003	4.0049e-003	4.0073e-003	4.7945e-003
	EQM	2.7140e-005	2.6642e-005	2.6952e-005	3.5159e-005
k=8	EM	1.3572e-004	7.3653e-005	-4.5095e-004	1.8425e-003
	EAM	4.6494e-003	4.6312e-003	4.6144e-003	5.0819e-003
	EQM	3.5366e-005	3.5096e-005	3.5448e-005	3.9480e-005
		Heures Travaill�ees			
Horizon	Erreur	MSF	MCI	MHP	VAR
k=1	EM	5.6227e-003	3.7762e-003	3.3083e-003	3.1938e-003
	EAM	5.6372e-003	4.3542e-003	4.3442e-003	4.1358e-003
	EQM	4.2274e-005	2.6791e-005	2.9111e-005	2.4250e-005
k=4	EM	1.1651e-002	9.1114e-003	9.9385e-003	1.0047e-002
	EAM	1.1708e-002	9.4077e-003	1.0547e-002	1.0489e-002
	EQM	1.7891e-004	1.2442e-004	1.6069e-004	1.6228e-004
k=8	EM	1.5904e-002	1.3994e-002	1.7323e-002	1.8535e-002
	EAM	1.6998e-002	1.5271e-002	1.8365e-002	1.9253e-002
	EQM	4.0410e-004	3.5139e-004	5.1415e-004	6.0026e-004

Au terme de cette section, il apparaît que les modèles structurels de taille réduites peuvent rivaliser avec les modèles VAR tant du point de vue de leur capacité à rendre compte de la vraisemblance des données, que de leur pouvoir prédictif. Il convient cependant de se garder de conclure à la domination des modèles structurels sur la modélisation VAR dans la mesure où les tests précédents concluent d'avantage à une propriété d'équivalence qu'à une stricte domination. La section suivante fournit un exemple supplémentaire de l'avantage des modèles structurels sur les modèles VAR.

5 Analyse structurelle

Au delà de la capacité des modèles à prévoir l'évolution future des variables modélisées, il est également intéressant de fournir une interprétation économique des résultats. Il convient alors de se tourner vers l'analyse structurelle du modèle. Le modèle structurel fournit par construction un outil permettant une telle interprétation. Dans les modèles proposés dans la présente étude, les comportements de l'ensemble des variables sont entièrement spécifiés et les modèles permettent l'analyse de l'ensemble des agrégats macroéconomiques dans un cadre unifié. En particulier, il est possible d'étudier la dynamique de chacun des agrégats — outre ceux sur lesquels est conduite l'estimation — suite à des chocs sur la productivité totale des facteurs et/ou sur les préférences des agents. Une telle analyse n'est pas directement possible dans le cadre de la modélisation VAR dans la mesure où les résidus des équations constitutives du VAR ne sont pas structurels. Une approche communément retenue dans la littérature consiste alors à réécrire les résidus comme une combinaison de chocs structurels orthogonaux qu'il convient alors d'identifier.

La représentation VAR de la dynamique du processus centré $\{Y_t\}_{t=1}^T$, Y_t est simplement donnée par

$$A(L)Y_t = \varepsilon_t,$$

où $A(L) = (I - A_1L - \dots - A_\ell L^\ell)$ est un polynôme en l'opérateur retard L , et ℓ désigne le nombre de retard du VAR. ε_t désigne le vecteur des innovations canoniques du processus. ε_t est de moyenne nulle et de matrice de variance covariance $E\{\varepsilon_t \varepsilon_t'\} = \Sigma$. L'économiste est alors intéressé non par les innovations canoniques ε_t du VAR mais par des innovations structurelles, η_t . Il est admis que les innovations canoniques s'écrivent comme une combinaison linéaire des innovations structurelles

$$\varepsilon_t = S\eta_t,$$

où S est une matrice non singulière. Etant donné que les données ne nous renseignent que sur les innovations canoniques, il convient de formuler certaines hypothèses sur η_t .

Il est commun d'imposer l'orthogonalité des chocs structurels ainsi qu'une condition de normalisation telle que $E\{\eta_t \eta_t'\} = I_{n_y}$. Ainsi, l'identification de la matrice S est assurée en résolvant

$$SS' = \Sigma$$

Cependant, Σ n'impose que $n_y(n_y + 1)/2$ contraintes et ne permet ainsi pas d'identifier les n_y^2 éléments de S . Il convient alors d'imposer certaines restrictions. Dans la mesure où le VAR comprend 2 variables, Σ impose 3 restrictions. Il est donc nécessaire d'introduire une restriction supplémentaire. A des fins de cohérences, le modèle VAR structurel est spécifié et identifié en conformité avec les modèles théoriques. Dans tous les modèles structurels, le choc technologique est supposé suivre une marche aléatoire avec dérive, et exerce ainsi un effet sur la dynamique de long terme de l'ensemble des variables soumises aux processus de croissance. En revanche le choc de préférences n'a qu'un effet transitoire. Nous suivons donc Blanchard et Quah (1989) et Galí (1999) et identifions le choc technologique comme le seul choc ayant un effet à long terme sur la productivité horaire. L'autre choc sera donc identifié au choc temporaire. En définissant $C(L) = A(L)^{-1}S$ et en considérant que Y_t est composé du taux de croissance de la productivité et des heures travaillées (dans l'ordre indiqué), l'identification du choc permanent impose que la matrice $C(1)$ soit triangulaire inférieure. Dans ce cas, $C(1)$ est fourni par la décomposition de Choleski de la matrice $A(1)^{-1}\Sigma A(1)^{-1'}$. $C(1)$ peut ainsi être obtenu simplement à partir de l'estimation du modèle VAR. La matrice de passage, S , entre résidus canoniques, ε_t , et chocs structurels, η_t , est alors donnée par $S = A(1)C(1)$. La forme moyenne mobile infinie structurelle est alors

$$Y_t = A(L)^{-1}A(1)C(1)\eta_t$$

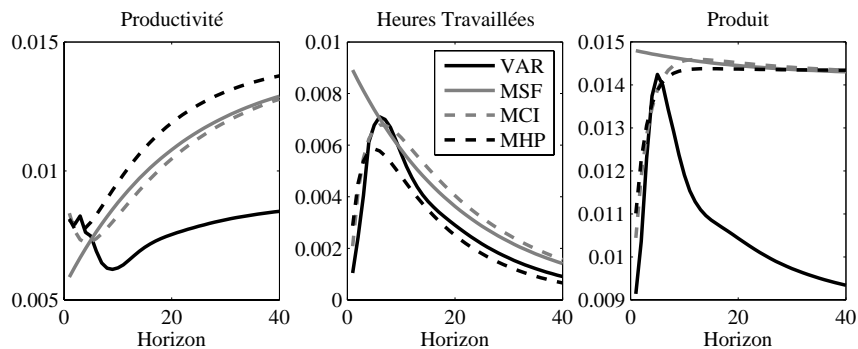
On en déduit alors les fonctions de réponse du modèle VAR aux chocs structurels.

Les figures 5 et 6 reportent les fonctions de réponse de la productivité en niveau, des heures travaillées et du produit par tête en niveau à un choc technologique et à un choc de préférences pour chacun des modèles.¹⁶ L'analyse des réponses montre que le modèle VAR et le modèle sans frictions (MSF dans le graphique) diffèrent largement du point de vue de la dynamique qu'ils engendrent. Plusieurs interprétations de ce résultat sont possibles. La première est que le schéma d'orthogonalisation des chocs dans le VAR ne permet pas d'identifier correctement les chocs technologiques et les chocs de préférences. Une seconde est que le modèle sans frictions n'est pas le bon modèle. L'analyse précédente a montré que cette dernière interprétation est la bonne dans la mesure où le modèle sans frictions ne présente pas de bonnes performances en prévision. En revanche, le modèle

¹⁶Etant donnée la définition des variables, la réponse du produit par tête est obtenue en additionnant celles des heures et de la productivité horaire.

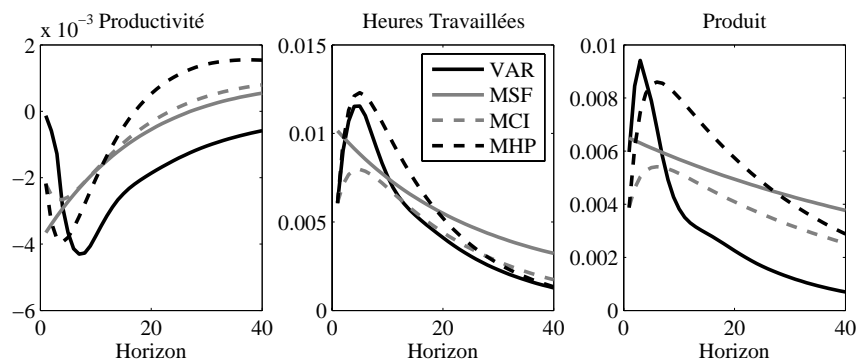
VAR et le modèle avec persistance des habitudes (MHP dans le graphique) engendrent une dynamique des heures très similaire. Ainsi, les fonctions de réponse aux deux chocs (*i*) sont de même ampleur tout au long de la trajectoire, (*ii*) présentent toutes deux un profil en cloche avec un pic à l'horizon 5 pour le modèle structurel et 6 pour le VAR suite à un choc technologique et 5 pour les deux modèles suite à un choc de préférences. Ceci vient confirmer l'analyse précédente qui concluait à l'équivalence entre les capacités prédictives du modèle VAR et du modèle structurel. Les deux modèles engendrent la même dynamique aux deux chocs. Ceci semble également suggérer que la décomposition à la Blanchard et Quah (1989) permet de bien identifier le choc technologique et le choc de préférences dans le modèle VAR, dans la mesure où le profil des réponses des heures à ces deux chocs sont semblables.

FIG. 5 – Fonctions de réponse — Choc technologique



Note : MSF : Modèle sans friction, MCI : Modèle à coûts d'ajustement sur l'investissement, MHP : Modèle avec persistance des habitudes.

FIG. 6 – Fonctions de réponse — Choc de préférences



Note : MSF : Modèle sans friction, MCI : Modèle à coûts d'ajustement sur l'investissement, MHP : Modèle avec persistance des habitudes.

L'analyse de la dynamique de la productivité vient cependant nuancer cette conclusion. En effet, si le profil des fonctions de réponse de la productivité est très similaire à court terme entre les deux modèles, il diffère sensiblement à plus long terme. Ainsi, l'effet de long terme d'un choc de productivité est beaucoup plus faible dans le VAR que dans le

modèle structurel quelle que soit la version considérée. Ce phénomène se retrouve alors dans la dynamique du produit. De la même façon, le modèle structurel engendre plus de persistance en réponse à un choc de préférences qu'à un choc transitoire dans le VAR. Ces résultats remettent en question la capacité du modèle VAR à identifier correctement tant les chocs technologiques que les chocs de préférences.

Afin d'illustrer ce phénomène, nous reconstruisons l'évolution conditionnelle des variables à chacun des chocs. Ceci revient à reconstruire, par exemple, la série de productivité sous l'hypothèse que seuls les chocs de productivité se sont réalisés. Dans le modèle VAR, ces séries sont obtenues en utilisant la représentation moyenne mobile infinie structurelle du modèle VAR. Etant donné le schéma d'orthogonalisation retenu, η_t est composé d'un choc technologique, $\eta_{z,t}$ et d'un choc de préférences $\eta_{\chi,t}$. La série conditionnelle au choc technologique (resp. choc de préférences) est alors simplement reconstruite en imposant $\eta_{\chi,t} = 0$ (resp. $\eta_{z,t} = 0$) pour tout t . Dans le modèle structurel, nous procédons de la même façon en utilisant la représentation espace-état. Les chocs technologiques et les chocs de préférences sont obtenus en appliquant un lissage de Kalman sur le modèle.

La corrélation entre l'innovation structurelle $\eta_{z,t}$ du modèle VAR et les chocs technologiques de chacun des modèles est remarquablement élevée. Elle s'élève à 0.93 quel que soit le modèle. Ceci est confirmé par l'analyse des corrélations entre les séries conditionnelles issues du VAR et du modèle structurel et illustré par les figures 7–9 qui reportent le graphe des séries conditionnelles obtenues dans le modèle structurel contre celles obtenues dans le modèle VAR.

L'analyse des graphiques montre que les séries de productivité et d'emploi reconstruites conditionnellement au seul choc technologique sont alignées le long de la droite à 45°. La

TAB. 7 – Corrélation VAR–Modèle structurel

Modèle	Productivité		Heures travaillées	
	Technologiques	Préférences	Technologiques	Préférences
Absence de frictions	0.94	-0.14	0.77	-0.15
Coût à l'ajustement	0.95	-0.17	0.94	-0.15
Persistance des habitudes	0.95	-0.27	0.83	-0.16

Note : Les p-values sont reportées entre crochets.

corrélation entre les séries issues du VAR et des modèles structurels est élevée. Ainsi cette corrélation s'élève à 0.95 pour l'évolution de la productivité conditionnelle au seul choc technologique, cette corrélation est un peu plus faible pour l'emploi dans la mesure où elle s'élève à 0.83 dans le modèle avec persistance des habitudes — modèle qui possède de bonnes propriétés en termes (i) de vraisemblance, (ii) de prévision et (iii) d'auto-corrélation des résidus. Ceci semble indiquer que le modèle VAR permet d'identifier au

FIG. 7 – Décomposition historique (Modèle sans Frictions réelles)

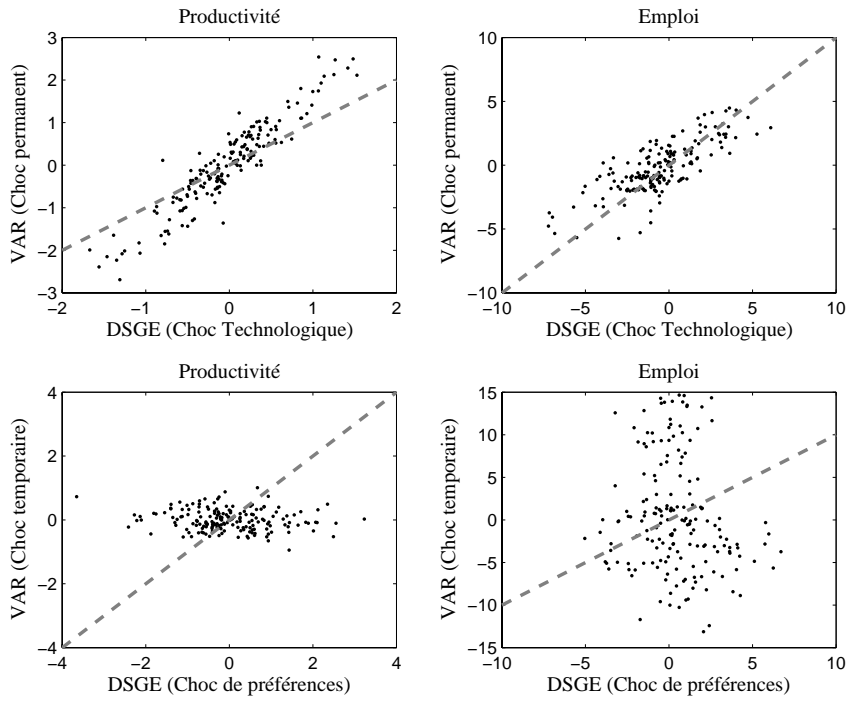


FIG. 8 – Décomposition historique (Modèle avec coûts d'ajustement)

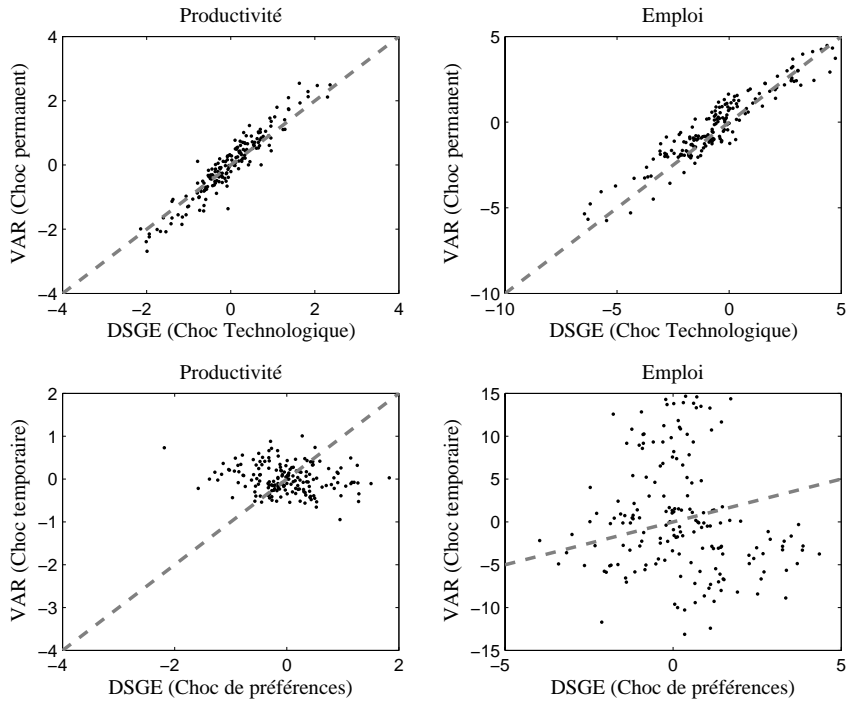
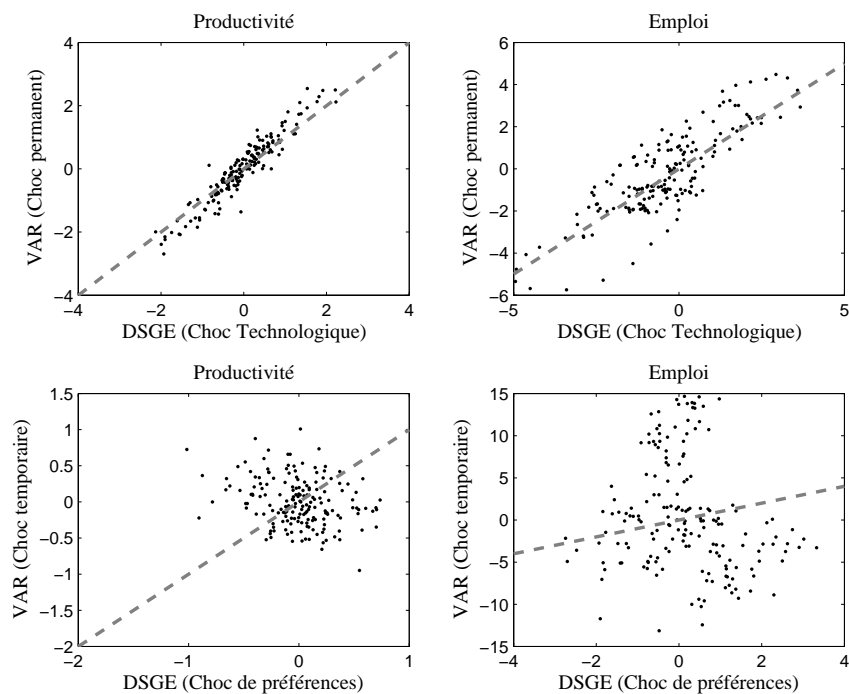


FIG. 9 – Décomposition historique (Modèle avec persistance des habitudes)



moins aussi bien le choc technologique que le modèle structurel.

En revanche, les séries de productivité et d’heures travaillées conditionnelles au choc de préférences présentent une corrélation négative entre le VAR et le modèle structurel. Ceci est illustré par les graphiques 7–9. Ainsi, le modèle VAR ne permet pas d’identifier correctement le choc de préférences.¹⁷ Il n’autorise donc pas une analyse structurelle de la composante transitoire des données. Ceci s’explique avant tout par la faible corrélation existante entre la composante transitoire du modèle VAR et le choc de préférences issu du modèle structurel (0.17). Ceci est une illustration de la difficulté des modèles VAR à identifier séparément une composante permanente et une composante transitoire très persistante. Ainsi, les chocs identifiés par le modèle VAR et les fonctions de réponses associées doivent être interprétés avec prudence.

6 Conclusion

Les travaux récents en macroéconométrie appliquée ont montré que des modèles DSGE de grande taille incluant des frictions réelles et nominales présentaient des qualités prédictives satisfaisantes, dominant celles des modèles VAR (voir Smets et Wouters (2003)). Cet

¹⁷Ce résultat ne contredit pas les résultats précédents (vraisemblances identiques des modèles VAR et DSGE, et effets des chocs technologiques identiques), puisque la contribution des chocs de préférences à la volatilité des séries est faible (voir Tableau 2).

article a montré que des modèles réels de taille réduite incluant un degré minimal de frictions réelles pouvaient concurrencer les modèles VAR pour la prévision de certaines variables agrégées (dans ce cas, la productivité horaire du travail et les heures travaillées). Nous montrons en particulier qu'une modélisation structurelle intégrant un phénomène de persistance des habitudes de loisir permet de prévoir au moins aussi bien qu'un modèle VAR incluant pourtant une dynamique plus fournie. Au delà, le modèle DSGE autorise une véritable interprétation structurelle des chocs et de la dynamique qu'ils engendrent, ce que ne permet pas toujours le modèle VAR.

Références

- BLANCHARD, O.J. ET D. QUAH, The Dynamic Effects of Aggregate Supply and Demand Disturbances, *The American Economic Review*, 1989, 79 (4), 655–673.
- BOUAKEZ, H. ET T. KANO, Learning-by-doing or habit formation?, *Review of Economic Dynamics*, 2006. Forthcoming.
- CHARI, V.V., P.J. KEHOE, ET E.R. MCGRATTAN, *A Critique of Structural VARs Using Real Business Cycle Theory*, Working Paper 631, Federal Reserve Bank of Minneapolis 2005.
- CHRISTIANO, L.J. ET J.D.M. FISHER, *Stock Market and Investment Goods Prices : Implications for Macroeconomics*, Working Paper 10031, NBER 2003.
- , M. EICHENBAUM, ET C.L. EVANS, Monetary Policy Shocks : What Have we Learned and to What End?, in M. Woodford et J. Taylor, éditeurs, *Handbook of Macroeconomics*, North-Holland, 1999, chapitre 3.
- , —, ET —, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy*, 2005, 113 (1), 1–45.
- , —, ET R. VIGFUSSON, *What Happens After a Technology Shock ?*, mimeo, Northwestern University 2004.
- COGLEY, T. ET J.M. NASON, Output Dynamics in Real–Business–Cycle Models, *The American Economic Review*, 1995, 85 (3), 492–511.
- DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, ET R. WOUTERS, On the Fit and Forecasting Performance of New Keynesian Models, *Journal of Business and Economic Statistics*, 2006. à paraître.
- EICHENBAUM, M. ET J.D.M. FISHER, Fiscal Policy in the Aftermath of 9/11, *Journal of Money, Credit and Banking*, 2005, 37 (1), 1–22.

- , L. HANSEN, ET K. SINGLETON, A Time Analysis of Representative Agent Models of Consumption and Leisure Choice under uncertainty, *Quarterly journal of Economics*, 1988, 79 (4), 733–48.
- GALÍ, J., Technology, Employment, and the Business Cycle : Do Technology Shocks Explain Aggregate Fluctuations?, *American Economic Review*, 1999, 89 (1), 249–271.
- ET P. RABANAL, *Technology Shocks and Aggregate Fluctuations : How Well Does the RBS Model Fit Postwar U.S. Data ?*, Working Paper 10636, NBER 2004.
- GOURIÉROUX, C. ET A. MONFORT, *Statistique et Modèles Econométriques*, Paris : Economica, 1989.
- ET — , *Séries Temporelles et Modèles dynamiques*, Paris : Economica, 1990.
- HAIRAULT, J.O., Présentation et Evaluation du Courant des Cycles Réels, *Economie et Prévision*, 1992, 106, 1–22.
- HAMILTON, J.D., *Times Series Analysis*, Princeton (NJ) : Princeton University Press, 1994.
- HANSEN, G., Indivisible Labor and the Business Cycles, *Journal of Monetary Economics*, 1985, 16 (3), 309–27.
- HANSEN, G.D., Technical Progress and Aggregate Fluctuations, *Journal of Economic Dynamics and Control*, 1997, 21, 1005–1023.
- KING, R ET S. REBELO, Resuscitating Real Business Cycles, in J. Taylor et M. Woodford, éditeurs, *Handbook of Macroeconomics*, Elsevier, 2000.
- KYDLAND, F.E. ET E.C. PRESCOTT, Time to Build and Aggregate Fluctuations, *Econometrica*, 1982, 50, 1345–1370.
- LUCAS, R., Econometric Policy Evaluation : A Critique, *Carnegie Rochester Conference Series on Public Policy*, 1976, 1, 19–46.
- ROTEMBERG, J. ET M. WOODFORD, An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy, in B. Bernanke et J. Rotemberg, éditeurs, *NBER Macroeconomics Annual*, Cambridge (MA) : MIT Press, 1997, pp. 297–346.
- SIMS, C., Macroeconomics and Reality, *econometrica*, 1980, 48 (1), 1–48.
- , Interpreting the Macroeconomic Time Series Facts : The Effects of Monetary Policy, *European Economic Review*, 1986, 36 (5), 975–1000.
- SMETS, F. ET R. WOUTERS, An estimated stochastic dynamic general equilibrium model of the euro area, *Journal of European Economic Association*, 2003, 1, 1123–1175.

A Evaluation de la log-vraisemblance

En appliquant l'opérateur espérance à l'équation (4), il vient simplement

$$X_{t|t-1} = M_X(\theta)X_{t-1|t-1}$$

De la même façon, la matrice de variance-covariance des erreurs de prévision peut être obtenue par

$$P_{t|t-1} = M_X(\theta)P_{t-1|t-1}M_X(\theta)' + M_E(\theta)\Sigma M_E(\theta)'$$

L'algorithme doit cependant être initialisé par $X_{0|0}$ et $P_{0|0}$. Dans la mesure où le modèle est log-linéarisé autour de l'état stationnaire, déterministe, nous supposons que l'état initial du système correspond à cet état stationnaire. Ainsi, $X_{0|0} = 0$ et $P_{0|0}$ est donné par la matrice de variance-covariance P^* à l'état stationnaire

$$P^* = M_X(\theta)P^*M_X(\theta)' + M_E(\theta)\Sigma M_E(\theta)'$$

Le calcul de $X_{t|t-1}$ et $P_{t|t-1}$ nécessite celui de $X_{t|t}$ et $P_{t|t}$. Ces deux dernière quantités sont obtenue par le filtre de Kalman (voir Hamilton (1994) pour plus de détails). En définissant la matrice de gain

$$K_t = P_{t|t-1}M_Y(\theta)(M_Y(\theta)P_{t|t-1}M_Y(\theta)')^{-1},$$

le meilleur prédicteur de X_t sachant l'information disponible au début de la période t ($X_{t|t}$) est alors donné par

$$X_{t|t} = X_{t|t-1} + K_t(Y_t - M_Y(\theta)X_{t|t-1})$$

et la matrice de variance covariance associée, $P_{t|t}$, par

$$P_{t|t} = P_{t|t-1} - K_tM_Y(\theta)P_{t|t-1}$$

La log-vraisemblance pour la période t est alors

$$\begin{aligned} \log(f(Y_t|\mathcal{Y}_{t-1}; \theta)) &= -\frac{n_y}{2} \log(2\pi) - \frac{1}{2} \log(|M_Y(\theta)P_{t|t-1}M_Y(\theta)'|) \\ &\quad - \frac{1}{2} (Y_t - M_Y(\theta)X_{t|t-1})' (M_Y(\theta)P_{t|t-1}M_Y(\theta)')^{-1} (Y_t - M_Y(\theta)X_{t|t-1}) \end{aligned}$$