

Discussion du papier
*Apports de Google Trends pour prévoir la
conjuncture française: des pistes limitées*
par C. Bortoli et S. Combes

Laurent Ferrara
Banque de France - DGEI
Service de macroéconomie internationale

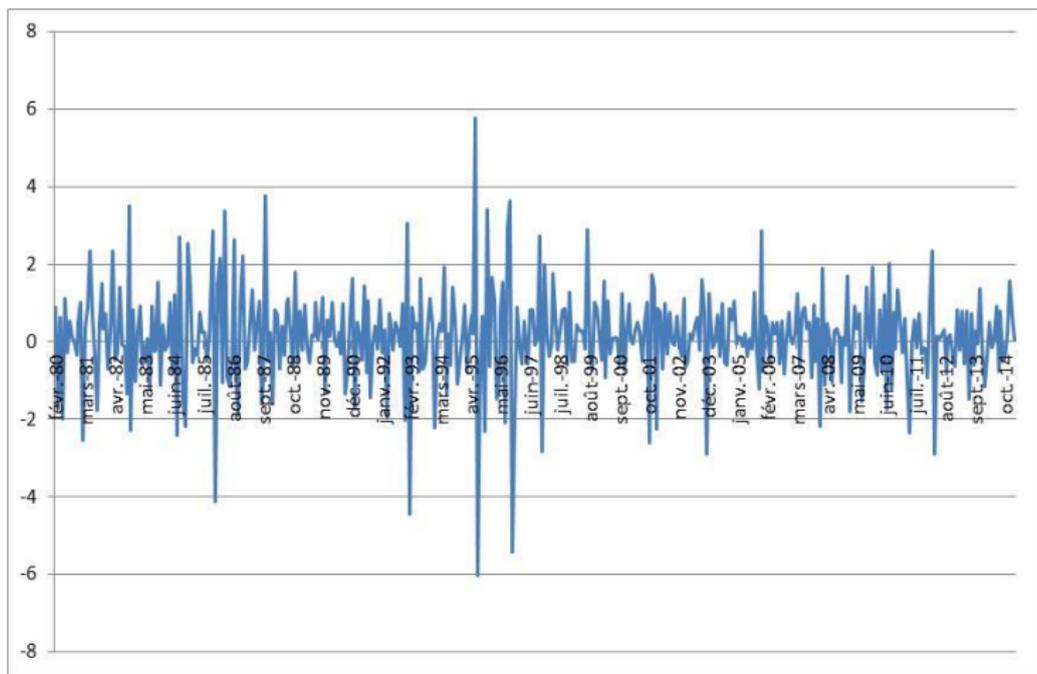
Séminaire Fourgeaud
8 avril 2015

Résumé

1. **Objet:** Préviation de la consommation des ménages avec de l'information provenant de Google Trends
2. **Données:** Croissance mensuelle de la conso des ménages prévue à 1 mois par des séries issues de Google Trends
3. **Méthodes:** Combinaison bayésienne de modèles ARDL (*Auto-Regressive Distributed Lags*)
4. **Résultats:** Pas de gain sur la conso agrégée, mais sur certains postes: habillement, équipement du logement, alimentaire
5. **Conclusions:** mitigées, notamment sur la qualité des données Google

Article très clair, bien écrit, sujet très intéressant (extension des travaux de Hal Varian).

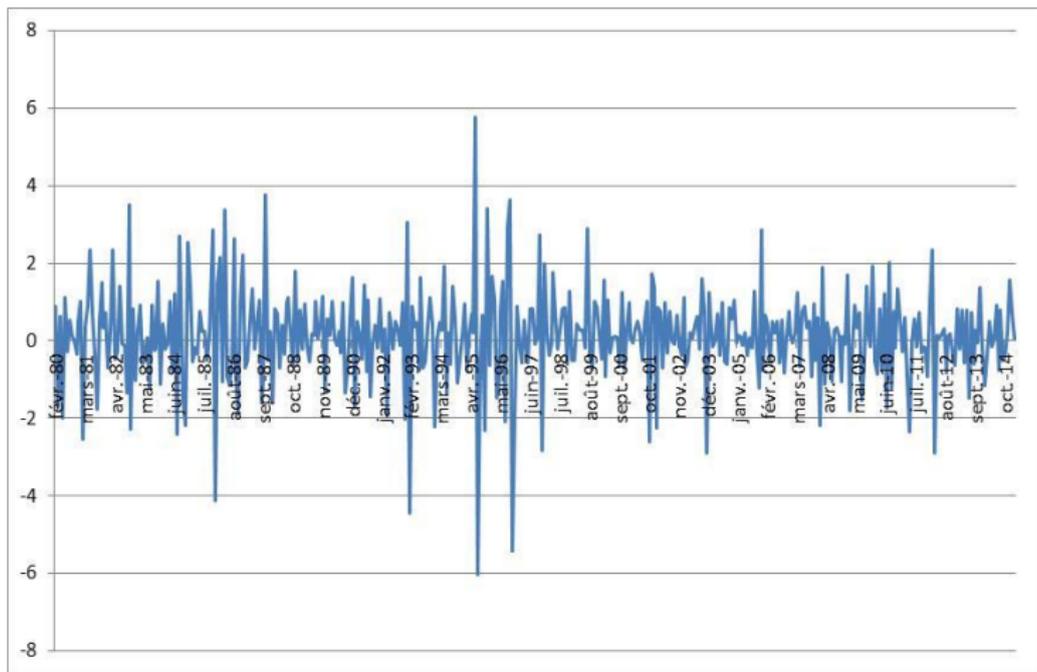
Quizz : quelle est cette série?



Quizz : quelle est cette série?

1. Faible dynamique $AR(p)$: $p = 2$ ou $p = 3$ selon la période
2. Forte hétéroscédasticité : Un processus GARCH(1,1) passe tous les tests
3. Non-Gaussianité : Test de Jarque-Bera rejette l'hypothèse nulle à cause d'une forte leptokurticité. Un Student-GARCH (avec 4 dl) semble ajuster au mieux les données.
4. Série difficile à prévoir : un modèle AR(2) estimé sur la période 2004-2011 n'améliore pas significativement les prévisions à 1 mois obtenues en utilisant la moyenne des valeurs passées sur la période 2012-2014
5. Cette série pourrait très bien être un rendement mensuel d'un actif financier

Croissance mensuelle de conso des ménages en biens

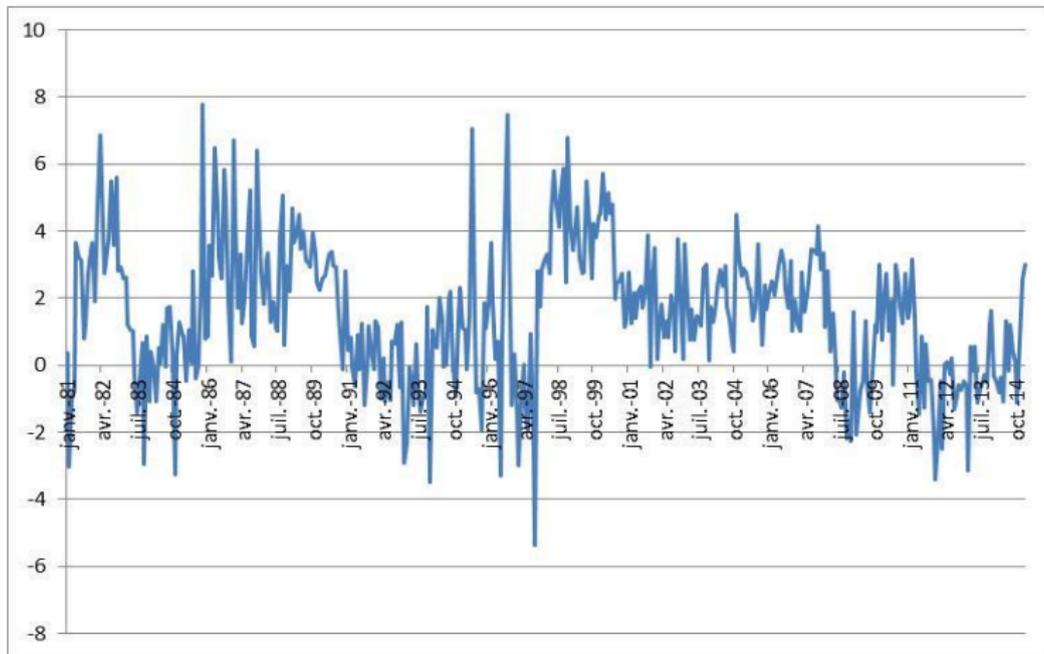


Les données

Variable à prévoir:

- ▶ Très forte volatilité du taux de croissance mensuel, peu de signal, série très difficile à prévoir
- ▶ Papier de référence de Vosen et Schmidt (JoF, 2011) compare les données Google Trends et les données d'enquête pour prévoir le **glissement annuel** de la conso US (voir figure suivante)
- ▶ Pourquoi ne pas essayer de prévoir/*nowcaster* la conso des ménages issue des comptes trimestriels ? Dans ce cadre cela serait intéressant d'évaluer l'apport de Google Trends au cours de la période de prévision par comparaison aux enquêtes de conjoncture et données *dures*.

Le glissement annuel de la conso des ménages en biens



Les données

Variables explicatives:

- ▶ Changements structurels dans l'utilisation de Google
- ▶ Le problème du biais d'échantillonnage est majeur : Quelles solutions possibles ?
 - ▶ Travail statistique à effectuer à la source par Google
 - ▶ Réduire ce biais en augmentant le nombre de tirages et en moyennant
 - ▶ Autres?
- ▶ Comment traiter la saisonnalité ? Peu de recul historique sur les données explicatives (depuis 2004). Faut-il corriger les séries hebdomadaires ou mensuelles ? Quelle méthode utiliser?

Les modèles et enjeux en termes de modélisation

1/ Grand nombre de données explicatives

- ▶ **Combinaison bayésienne de modèles** (Raftery et al., 97)
- ▶ Modèles à facteurs dynamiques (Forni et al., 2000, Giannone et al, 2008, Stock et Watson, 2002, une revue dans Barhoumi et al. EP 13)) et extensions e.g.: *Targeted Dynamic Factor Model*
- ▶ Régressions pénalisées de type LASSO (Tibshirani, 96) ou *Ridge regressions*
- ▶ *Bayesian shrinkage* dans des VAR à grande dimension (Banbura, Giannone, Reichlin, 10)
- ▶ Algorithmes de sélection de variables de type GETS ou Autometrics (Hendry et co-auteurs) ou approche *Spike and Slab*

Les modèles et enjeux en termes de modélisation

2/ Variables de périodicités différentes: mensuelles ou trimestrielles (à gauche) vs hebdomadaires (à droite)

- ▶ Solution simple: agréger à la fréquence la plus basse et estimer des modèles ARDL
- ▶ Créer des séries à basse fréquence à partir des variables à haute fréquence (approche *blocking* ou U-MIDAS)
- ▶ Utilisation des modèles à mélange de fréquences : MIDAS (Ghysels et différents co-auteurs). Nombreuses extensions des modèles. Nombreuses applications en prévision macro et en *nowcasting*
- ▶ Dans un VAR de faible dimension, considérer la fréquence basse avec des données manquantes et filtrer avec Kalman (Mariano et Murasawa, JAE 03)

La régression MIDAS

- ▶ Le cadre **MIDAS** fournit une approche parcimonieuse pour gérer une régression multi-fréquence (Ghysels et al., 07)
- ▶ Equation avec y_t à expliquer et x_t^k explicative à plus haute fréquence :

$$y_t = \alpha + \beta \text{midas}^K(\theta, L) x_t^k + \varepsilon_t \quad (1)$$

où $\text{midas}^K(\cdot)$ permet de lisser les K dernières valeurs de x_t^k :

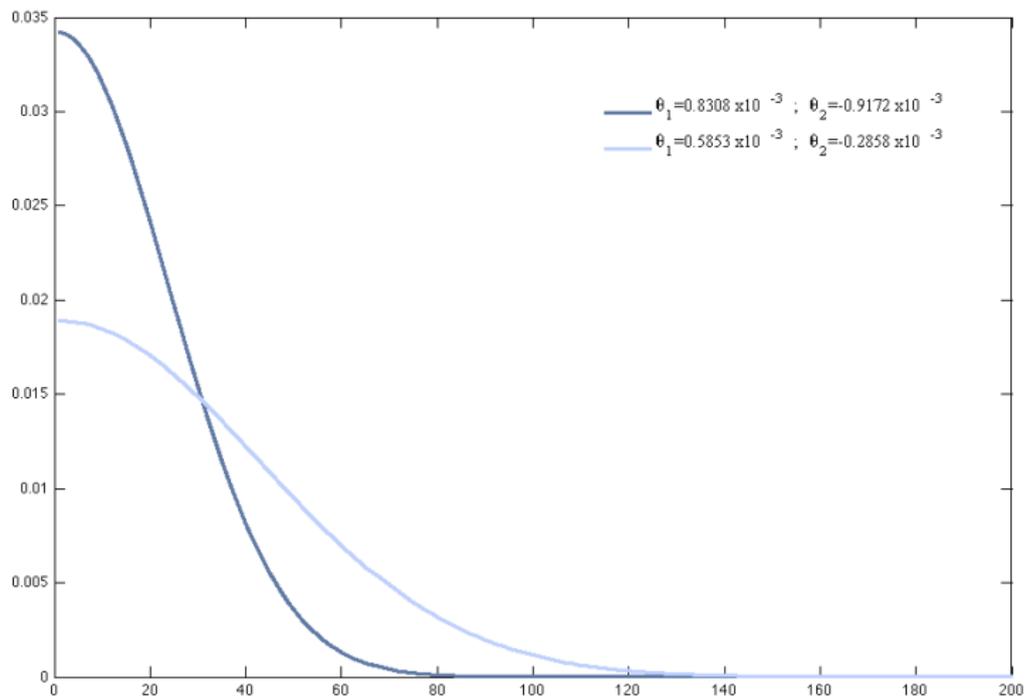
$$\text{midas}^K(\theta, L) := \sum_{k=0}^{K-1} \frac{f_K(k, \theta)}{\sum_{l=0}^{K-1} f_K(l, \theta)} L^k$$

avec par exemple:

$$f_K(k, \theta) = \exp(\theta_1 k + \theta_2 k^2)$$

La régression MIDAS

La fonction de poids



Les modèles multi-fréquence avec un grand nombre de variables: de nombreux travaux récents

- ▶ *Factor-Augmented MIDAS* (Marcellino et Schumacher, OBES 11, ou Ferrara et Marsilli, WP BdF 14)
- ▶ *Mixed Frequency VAR* avec changements de régimes (Faroni, Guérin, Marcellino, CEPR, 14)
- ▶ Marsilli (14) compare LASSO, sélection bayésienne de variables (approche *spike and slab*) et modèles à facteurs dynamiques dans un cadre MIDAS. Met en évidence que le poids de chaque variable évolue au cours du temps.
- ▶ Siliverstovs (15) estime un MIDASSO (U-MIDAS avec LASSO)

Quelques questions relatives au papier

- ▶ Plus de détails sur la mise en oeuvre de la modélisation seraient nécessaires dans le papier. Horizon ? Exercise récursif ou roulant ?
- ▶ Pourquoi privilégier la combinaison bayésienne ? Au moins une autre approche aurait été intéressante à mettre en oeuvre (à expliciter par exemple dans l'encadré 3).
- ▶ Benchmark de Kitchen et Monaco (03): n prévisions pour n variables explicatives
- ▶ les L modèles sélectionnés à l'aide de la *Marginal Likelihood* ou *Predictive likelihood* ?
- ▶ Aspect important: le poids de chaque variable explicative peut évoluer dans le temps. De quelle manière ? Lié au cycle économique?
- ▶ Remarque: présenter des ratios de RMSE, plus faciles à interpréter

Restons optimistes : les pistes à explorer

- ▶ Introduction de **non-linéarités** dans les modèles (lié à la variation au cours du temps du poids explicatifs de certaines variables).
D'Amuri et Marcucci (WP Bdl, 2012) : prévision du taux de chômage US.
Koop et Korobilis (2012) : *Dynamic model averaging* avec des modèles de régressions dont les paramètres dépendent du temps.
- ▶ Utilisation en prévision et / ou détection en temps réel des **points de retournement des cycles**. Voir par exemple Billio et al. (12) dans le cadre d'une combinaison bayésienne.

Restons optimistes : les pistes à explorer

- ▶ L'information Google n'est pas de grande qualité mais peut être utile dans des **cas où le praticien n'a pas d'information**.

Peu de références dans la littérature macro car l'information est assez bonne, sauf peut-être certain pays (application au Chili de Carriere-Swallow et Labbé, 10) ou certains secteurs (tourisme).

Exemple de la mesure de l'incertitude à la Bloom (09).

- ▶ **Exercice intéressant:** Anticiper la conso des ménage issue des comptes trimestriels avec comme variables explicatives les données Google hebdomadaires et les données d'enquêtes et *dures* mensuelles dans le cadre d'une régression MIDAS multi-fréquence (cf. Ferrara et al., Eco Mod 14).

Evaluer l'apport séquentiel au cours de l'horizon de chaque type de variables : Google /enquêtes/ donnée *dures*

Intuition: plus l'horizon est loin, plus l'information est pauvre, plus les variables Google peuvent avoir une capacité prévisionnelle importante

Conclusions

1. Un champ encore à explorer en termes d'applications (prix, marchés financiers ...) et de méthodes (modèles économétriques récents).
2. Pour certains secteurs (tourisme, santé) ou pays (émergents et en développement) pour lesquels la qualité et la quantité des données est moindre, le gain en termes de prévision à court terme / *nowcasting* peut être non-négligeable
3. Une piste intéressante : évaluer l'apport à l'analyse conjoncturelle des données Google Trends en termes d'anticipation des points de retournements des cycles

Quelques références

- ▶ Banbura, M., D. Giannone, L. Reichlin, 2010, Large Bayesian vector auto regressions, *Journal of Applied Econometrics*, 25(1), 71-92.
- ▶ Barhoumi, K. O. Darné and L. Ferrara (2013), Une revue de la littérature des modèles à facteurs dynamiques, *Economie et Prévision* No. 201.
- ▶ Carriere-Swallow Y. and F. Labbé, 2010, Nowcasting with Google Trends in an emerging market, *Banque Centrale du Chili*, WP No. 588.
- ▶ D'Amuri F. and J. Marcucci, 2012, Google it! Forecasting the US employment rate with a Google Job search index, *Banca d'Italia*, WP No. 891.
- ▶ Ferrara L. , C. Marsilli, 2014, Nowcasting global economic growth: A factor-augmented mixed-frequency approach, *Working papers* 515, Banque de France.
- ▶ Ferrara, L., Marsilli, C. and Ortega, J-P., 2014, Forecasting growth during the Great Recession: is financial volatility the missing ingredient?, *Economic Modelling*, vol. 36(C), pages 44-50.
- ▶ Forni, C., Guérin, P., Marcellino, M. 2014, Markov-Switching Mixed-Frequency VAR Models, *CEPR Discussion Papers* 9815.
- ▶ Ghysels, E., Sinko et Valkanov (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26, 53-90.
- ▶ Giannone D., Reichlin L. et Small D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55, 665-676.
- ▶ Kitchen J. and R. Monaco, 2003, Real-time forecasting in practice, *Business Economics*.
- ▶ Marcellino M. et Schumacher, C. (2010). Factor-MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72, 518-550.
- ▶ Mariano R. et Murasawa Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*. 18, 427-443.
- ▶ Marsilli C., 2014, Variable Selection in Predictive MIDAS Models, *Working papers* 520, Banque de France.
- ▶ Stock J. et Watson M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147-162.
- ▶ Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1), 267-288.
- ▶ Vosen S. and Schmidt (2011), Forecasting private consumption: Survey-based indicators vs. Google Trends, *Journal of Forecasting*, 30, 6, 565-578.