

Science des données?

Dalibor Stevanovic

Professeur, Sciences économiques, ESG UQAM

Co-titulaire, Chaire en macroéconomie et prévisions

Fellow CIRANO et Responsable du Pôle en modélisation

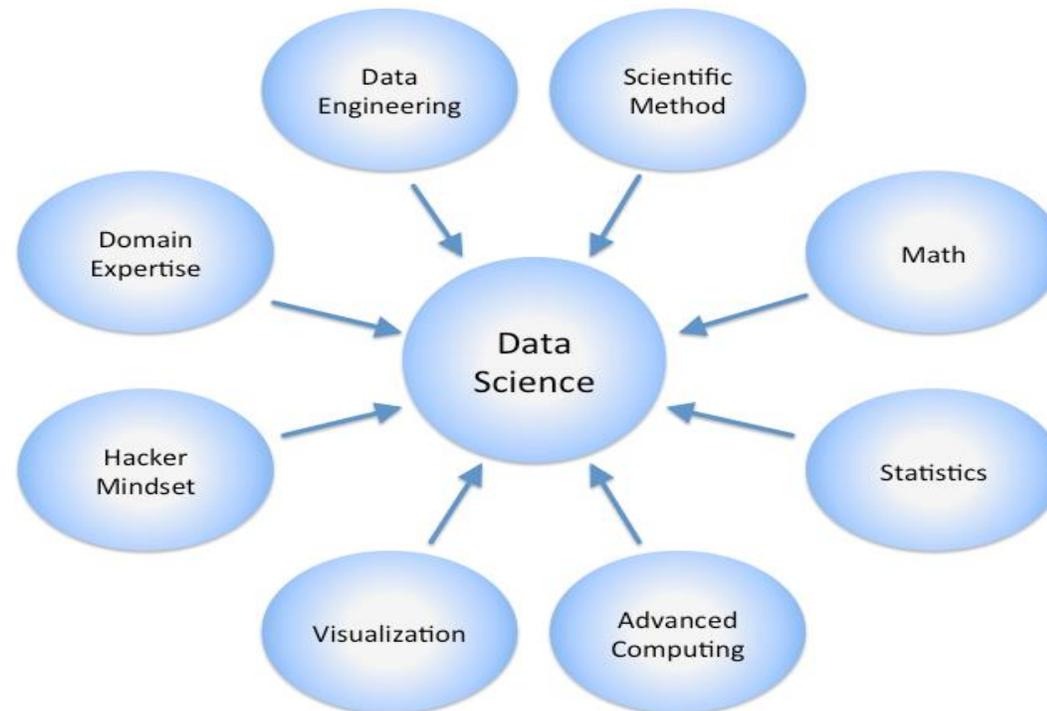
Colloque CIRANO: Relever les défis des chaînes
d'approvisionnement

Qu'est-ce que la *Science des données*?

- Définition (Amazon):
 - l'étude des données afin d'en extraire des informations significatives pour les entreprises... approche pluridisciplinaire issue des domaines des mathématiques, des statistiques, de l'intelligence artificielle et du génie informatique, en vue d'analyser de grands volumes de données. **Cette analyse aide les scientifiques des données à poser des questions et à y répondre, comme Que s'est-il passé, Pourquoi cela s'est-il passé, Que va-t-il se passer et Que peut-on faire avec des résultats.**
- Définition (Université d'Ottawa):
 - l'étude, l'application et l'élaboration de méthodes pour **tirer des leçons de ces données** afin de **comprendre** et **prévoir** les stratégies de gestion, les produits, les services, les campagnes publicitaires, la santé et sécurité publiques et bien plus... nécessitant la combinaison de mathématiques, informatiques et statistiques.

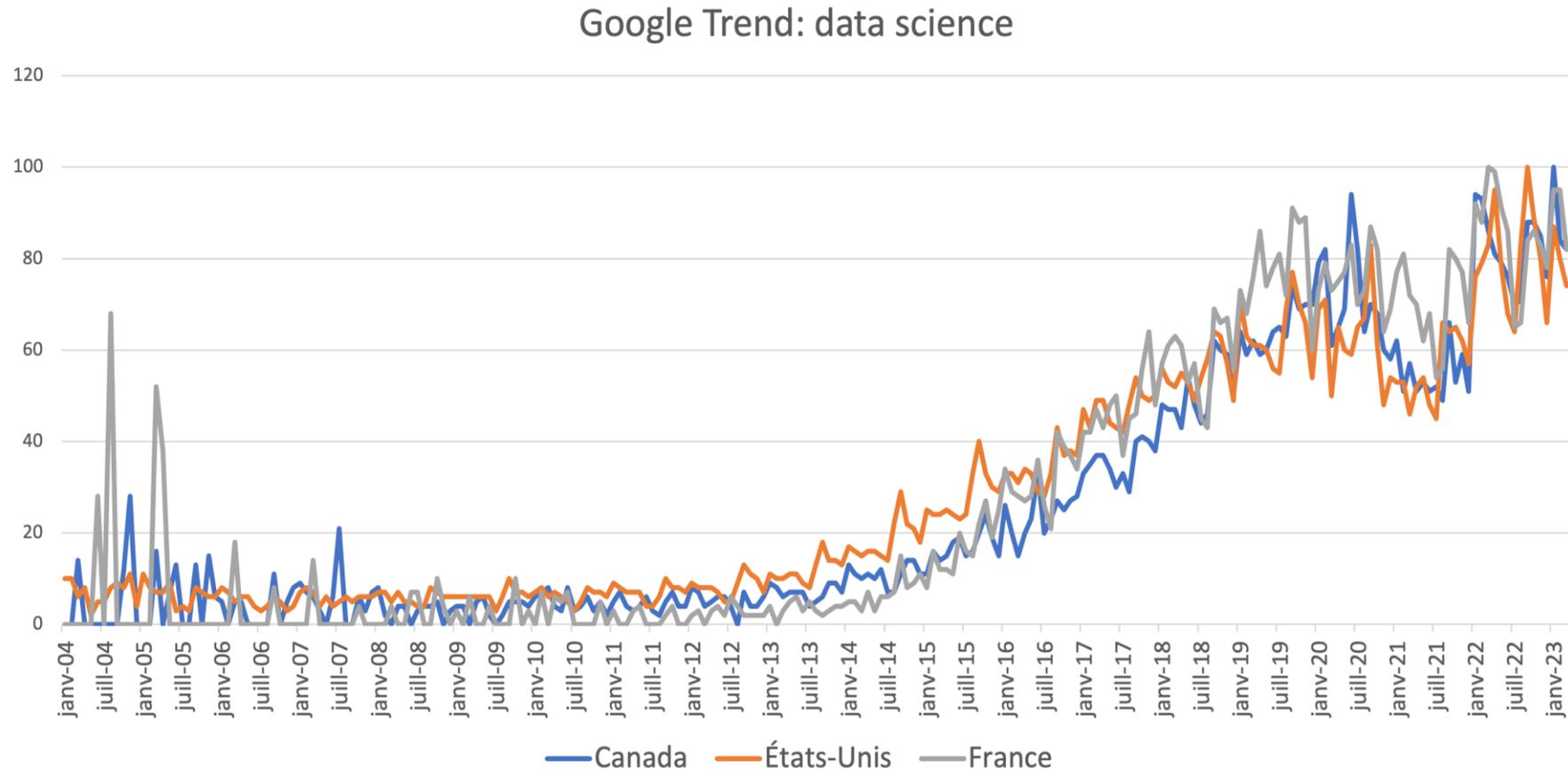
Qu'est-ce que la *Science des données*?

- Définition (Wikipedia):
 - l'étude de l'extraction automatisée de connaissance à partir de grands ensembles de données



Qu'est-ce que la *Science des données*?

- Pas le premier à se poser la question!



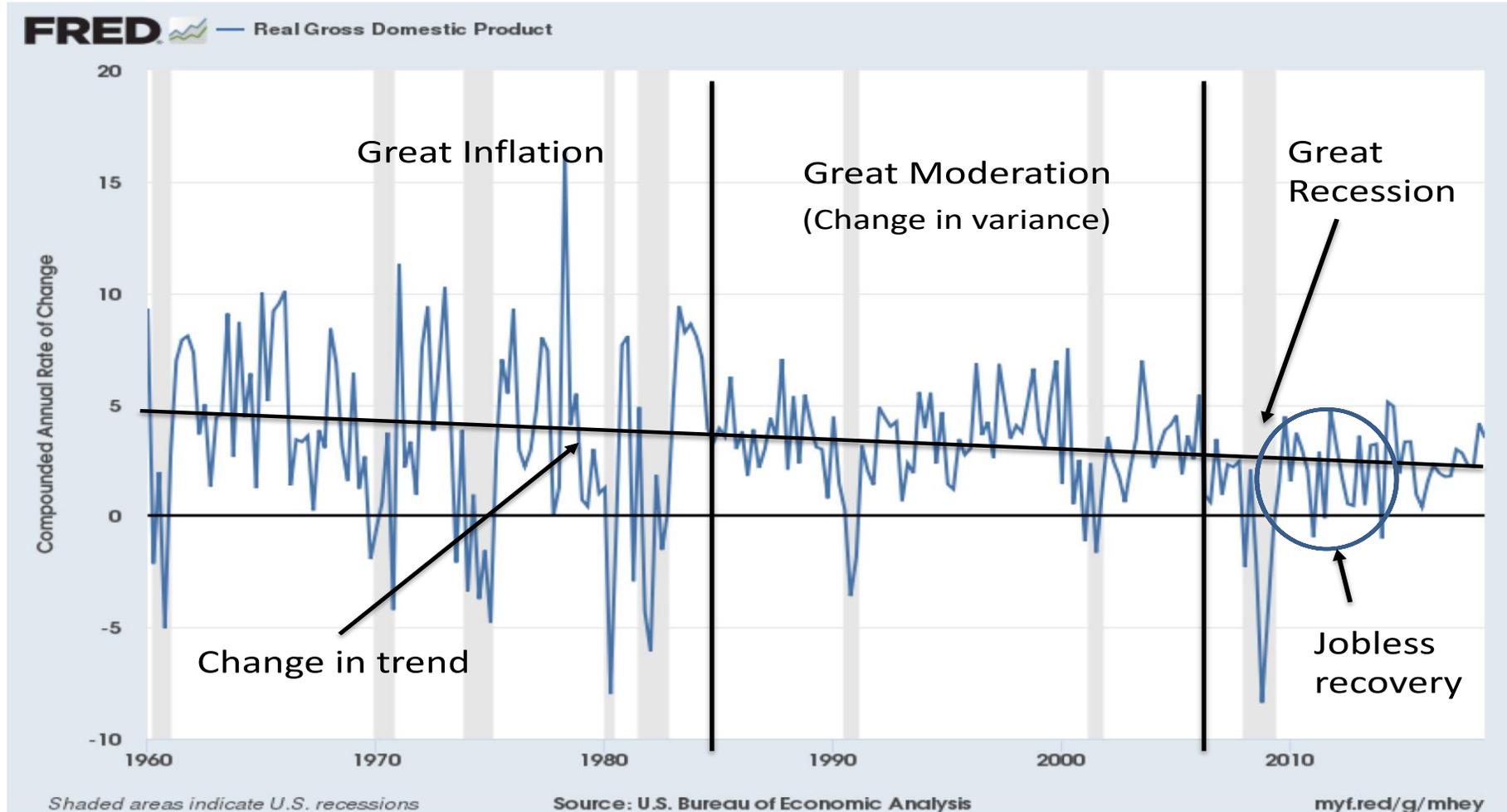
Qu'est-ce que la *Science des données*?

- Dans le contexte économique, c'est plutôt de l'économétrie appliquée:
 - Spécifier et estimer les modèles empiriques dans le but de:
 - *Documenter* les faits
 - *Prévoir* les phénomènes
 - *Essayer* d'expliquer (comprendre)
 - *Analyser* les impacts (pas facile)
 - Différences:
 - Modélisation nonlinéaire / nonparamétrique prône au surajustement (overfitting)
 - Tend à attribuer un pouvoir explicatif (prédictif) à l'aléatoire
 - Nécessite beaucoup de données et de capacité de calcul
 - Boite noire
 - Utile surtout pour la prévision
 - Ignore souvent l'incertitude (inférence)
 - Peu de progrès encore dans l'analyse causale

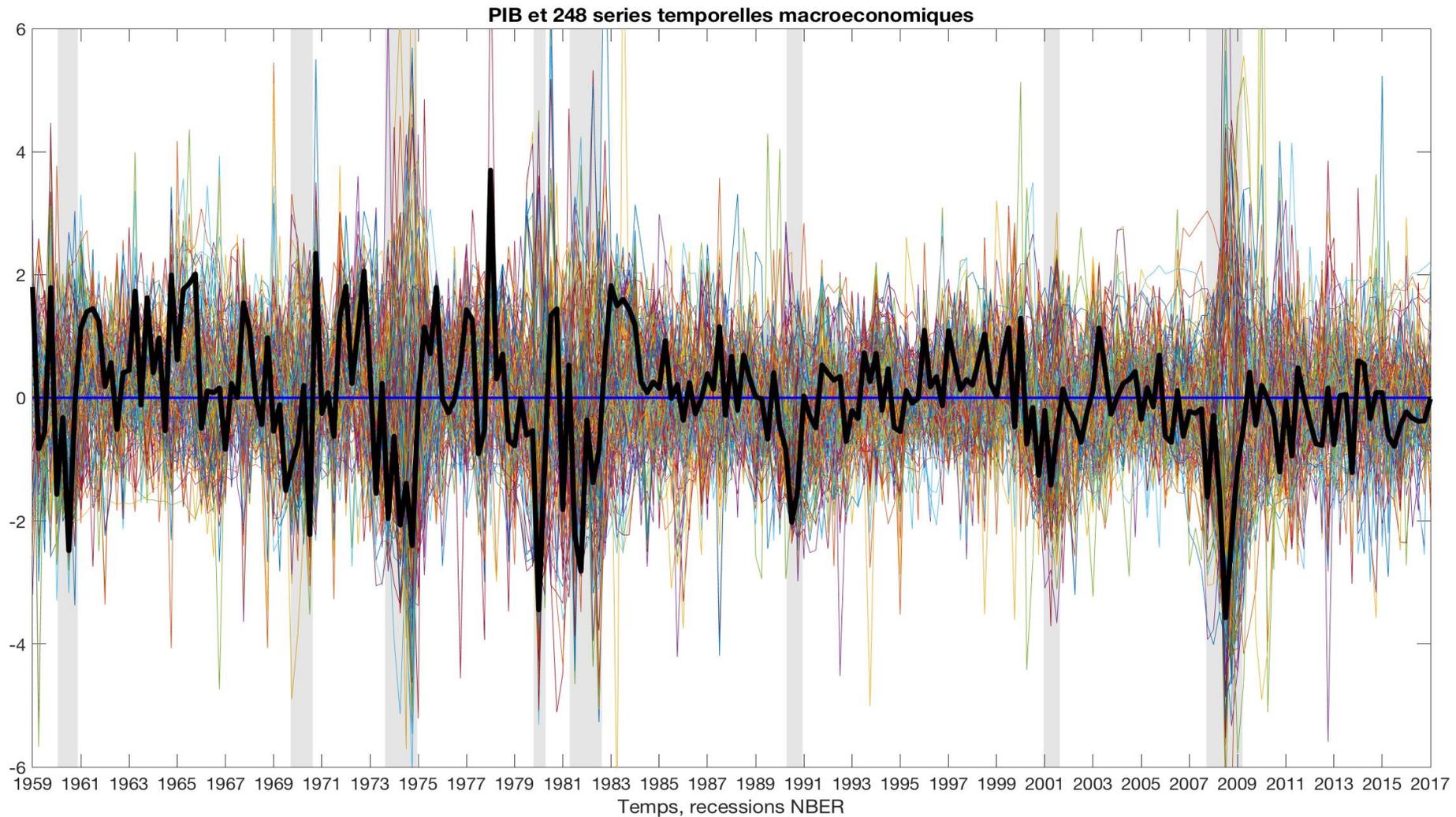
Utilité pour le corridor Saint-Laurent?

- Prédiction de la demande
 - Nationale: PIB
 - Internationale: PIB des partenaires, conditions économiques globales, flux commerciaux
- Prédiction de l'état du cycle économique
 - Récession?
- Analyse d'impact
 - Effets des chocs macroéconomiques sur
 - la demande
 - le commerce

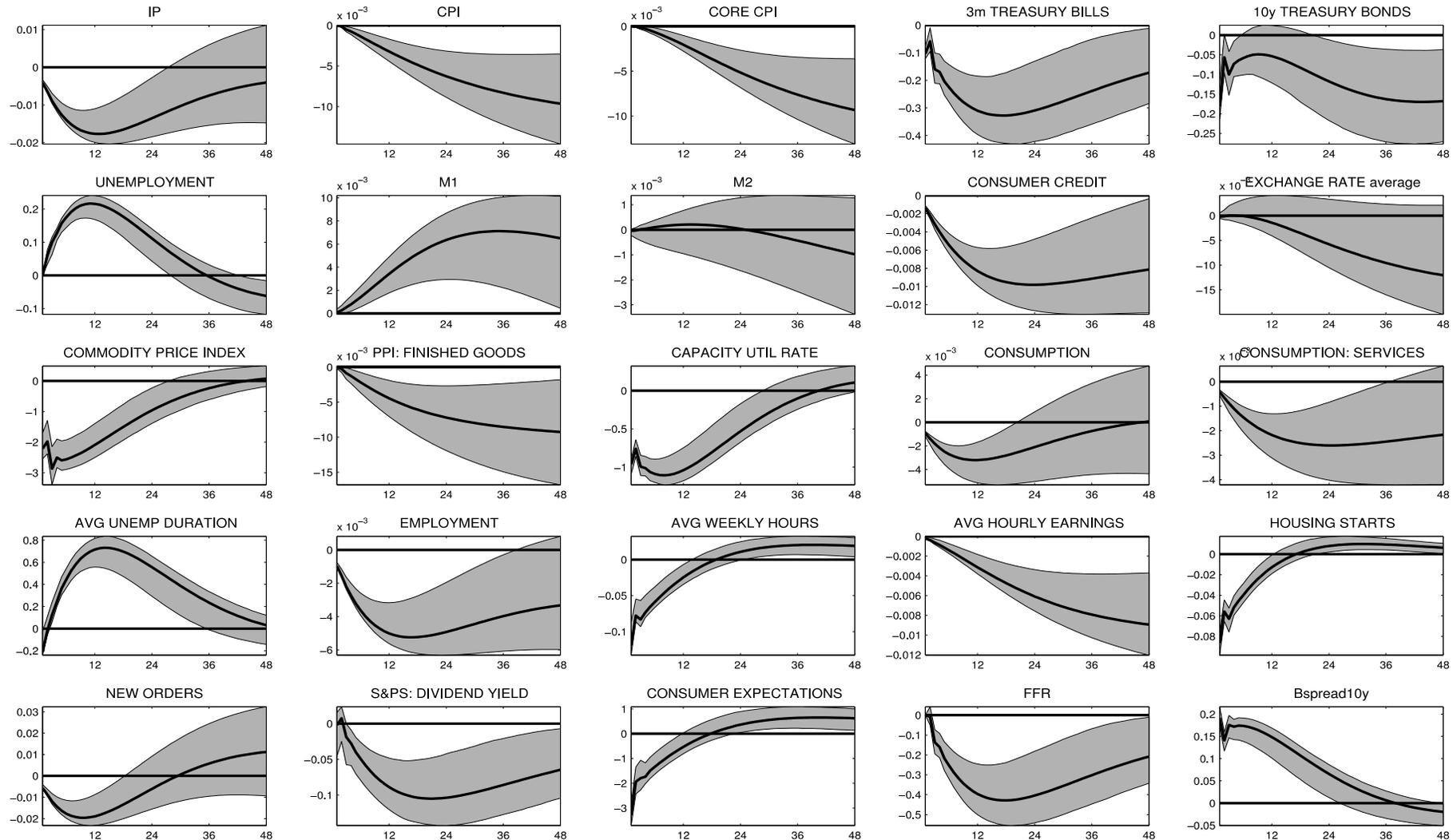
Prévoir le PIB des États-Unis...



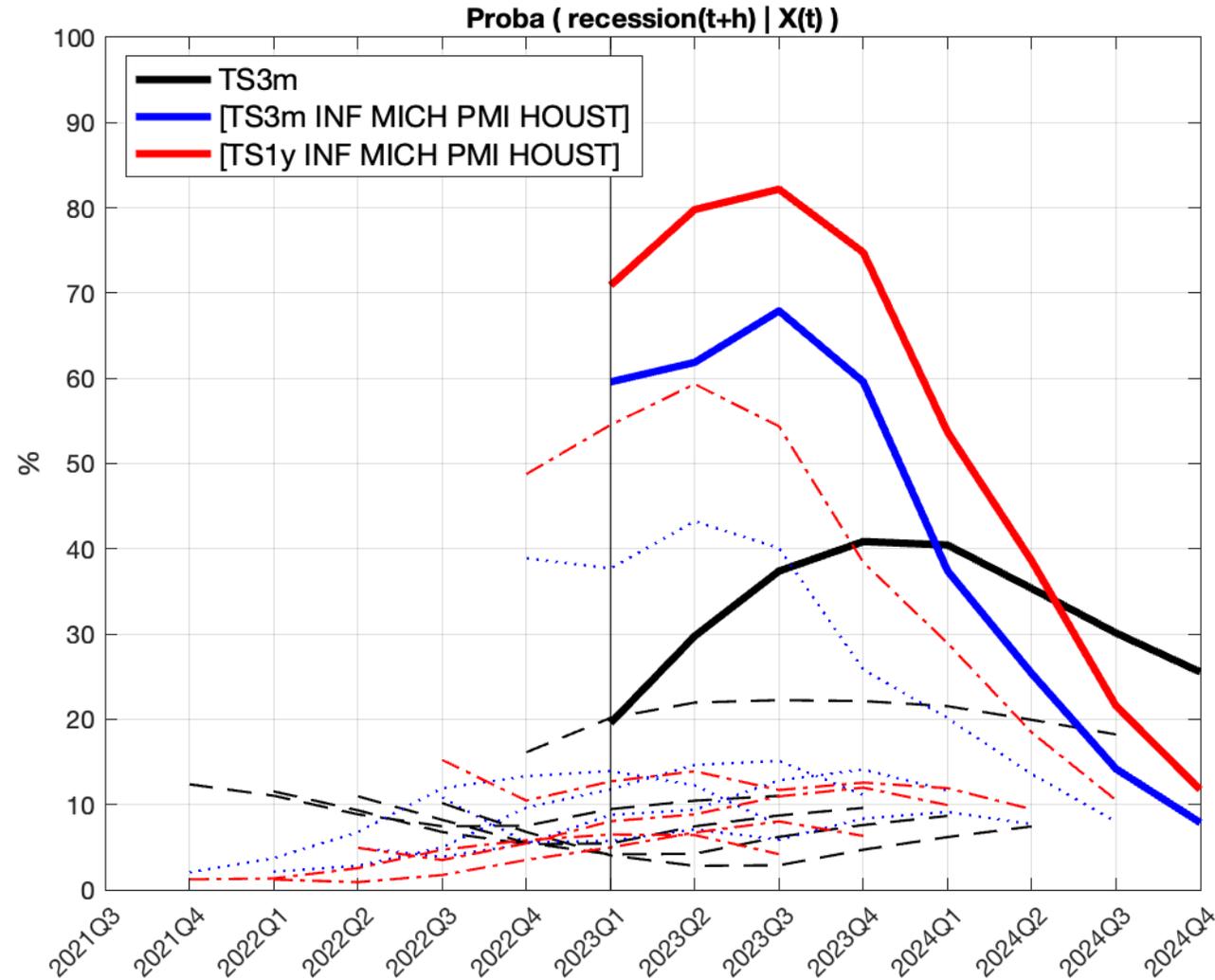
...avec beaucoup de données? (Goulet C., Leroux, Stevanovic et Surprenant, 2022)



Estimer les effets des chocs financiers (Boivin, Giannoni et Stevanovic, 2020)



Probabilité de récession? (Chaire en macroéconomie et prévisions)



Conclusions

- Science des données a produit des outils puissants permettant
 - Analyse des grandes bases de données
 - Amélioration des prévisions
 - Création et exploration des données non standard
- Angles morts
 - Boîtes noires
 - Nécessite une main d'œuvre hautement qualifiée pour ne pas tomber dans les pièges
 - Besoin d'économie ou d'autres sciences pour discipliner l'analyse
 - Nécessaire pour identifier la problématique et formuler la question que la SD peut adresser
 - Apprendre sur les mécanismes économiques à partir de la SD
- Potentiel pour le corridor
 - Meilleures prévisions → meilleure gestion
 - Meilleure modélisation → meilleure évaluation des risques
- Transfert des connaissances (\leftrightarrow) est primordial
 - CIRANO fournit un lieu d'échange essentiel