Le . do « first\_stage » commence par conserver les variables utiles, puis crée les variables utilisés lors des régressions :

- ratios de droit de douane (variable habituellement utilisée dans la littérature : droits de douane réglés/valeur importée)

- 3 dépendantes différentes : log du prix\_fob, le différentiel du précédent, le différentiel du prix\_fob non log-transformé.

- 3 IV construites symétriquement : ln (1+ tariff share), d.ln(1+tariff\_share), dtariff\_share

- génère deux bases de données contenant toutes les deux des valeurs prédites issues de régressions de first stage :

**1. Régressions en panel**

Deux fois six régressions identiques (1 groupe pour air, 1 groupe pour vessel).

Pour chaque groupe de 6, trois paires de régressions.

Chacune de ces 3 paires correspond à l’une des définitions de variable dépendante présentées plus haut (log, d.log, delta sans log).

Pour chacune des trois variables dépendantes, deux régressions (un paire) :

1ère reg : reg prix fob sur tariff avec FE pays-secteur 3 digit et FE secteur 3digit-année. Cette régression donne lieu alors à une prédiction qui donne un prix-fob\_hat, basé sur la prédiction de l’IV et des FE (c’est très important pour la suite).

2nde reg : la même que la précédente, mais avec des FE country\_year en plus. Ca prend environ 1 à 2 heures pour cette reg, c’est pour vérifier que la précédente ne souffre pas excessivement de biais de variable omise. On est obligés de procéder comme cela car reghdfe ne supporte que 2 dimensions de FE, pas 3. Donc on ne peut pas faire de prédictions basées sur les 3 sortes de FE.

Les résultats des reg sont sortis sous forme de tableau latex et consignés dans le log « first\_stage\_panel », et les predicted dans la base « predictions\_FS\_panel.dta ». On n’y voit notamment que les tariffs sont souvent de bons prédicteurs des prix fob, un peu plus pour air que pour vessel, mais avec des réversions de signe curieuses dans le 2nd cas. A discuter.

**2/ Régressions en coupes instantanées («  yearly »)**

Même chose qu’en 1, sans évidemment la dimension année (donc seulement des FE country-secteur 3 digit), mais avec deux types de prédictions : l’une (xbd) inclut les FE estimés, l’autre les exclut (xb), autrement dit, produit une prédiction seulement basée sur l’IV (xb). Les résultats sont consignés dans le log « first\_stage\_yearly », et », et les predicted dans la base « FS\_yearly.dta ». Evidemment, ces predicted sont très différents. Instinctivement, je dirais qu’il faut pendre les predicted issus de xbd (cad, incluant les FE estimés), car on a des valeurs bcp plus en lignes avec les observés. Mais le souci est que, pour des pbs de multico, il est régulier que des FE ne soient pas estimés, et qu’en conséquence, la prédiction donne des valeurs manquantes. A discuter également.