



Tutoriel Avionic

François Limousin

7 juillet 2020

1 Objectif

Le modèle Avionic (Analyse Variantielle Input/Output Nationale, Importée et en Contenus) développé par le département des comptes nationaux de l'Insee est ouvert en accès libre sur Github ici : <https://github.com/InseeFr/Avionic>. Le but de ce document est de présenter les différents programmes mis à disposition sur la plateforme et leur articulation afin de permettre à chacun d'utiliser le modèle Avionic. Ce modèle permet trois types de travaux :

- Calculer les contenus en importations, en valeur ajoutée, en emploi de la demande finale.
- Simuler un choc de demande et en voir les effets.
- Simuler une variation de prix de la production et en voir les effets.

Les calculs peuvent être réalisés sur les données françaises. Ils peuvent aussi être réalisés sur des bases de données internationales, à des fins de comparaisons entre pays. Il est aussi possible de produire des résultats par catégories de ménages ainsi qu'à différents niveaux de nomenclature. La méthodologie du modèle est expliquée dans un document de travail téléchargeable ici : <https://www.insee.fr/fr/statistiques/3970824>. Sa lecture est fortement conseillée avant d'utiliser les programmes mis en ligne.

2 Contenu d'Avionic

2.1 Vue d'ensemble des programmes disponibles

Onze programmes sont mis à disposition sur la plateforme. Le programme `tuto_avionic` est le programme central qui permet d'utiliser le modèle Avionic. Tous les autres programmes sont transparents pour l'utilisateur et seront appelés automatiquement lors du lancement du programme `tuto_avionic`. Parmi les 10 programmes appelés automatiquement, on distingue ceux qui servent au lancement du modèle proprement dit, sur données françaises ou internationales (programmes intitulés "`contenus_`" ou "`variation_`" des programmes qui permettent ensuite de présenter les résultats du modèle sous d'autres angles (niveau de nomenclature ou par catégorie de ménages : programmes intitulés "`Avionic_agregations`" et "`Avionic CMpC`" et des autres programmes qui servent pour des calculs intermédiaires ou des mise en forme de données.

- `tuto_avionic` : programme à lancer et à suivre pour mettre en œuvre les calculs et simulations d'avionnic. Ce document décrit les sorties de ce programme pas à pas.
- `contenus.R` : permet de calculer les contenus en importations, en valeur ajoutée et en emploi de la demande finale en France.
- `contenus_international.R` : permet de calculer les contenus en importations et en valeur ajoutée de la demande finale avec les ventilations des bases inter-pays (international).
- `variation_df.R` : permet de simuler les effets d'un choc exogène de demande finale.
- `variation_df_international.R` : permet de simuler les effets d'un choc exogène de demande finale par pays.
- `variation_prix.R` : permet de simuler la variation du prix de production en fonction d'un choc exogène de prix.
- `variation_prix_international.R` : permet de simuler la variation du prix de production en fonction d'un choc exogène de prix par pays.
- `Avionic_agregations.R` : permet d'agréger à différents niveaux de nomenclatures les données.
- `Avionic_CmpC.R` : permet d'appliquer les résultats d'Avionic au compte des ménages par catégories.
- `decoupage_international.R` : permet de découper les données par pays et d'utiliser ensuite les bases inter-pays.
- `fonctions.R` : crée des fonctions de calculs intermédiaires qui seront utilisées dans d'autres programmes.

2.2 Environnement de travail

Les programmes fournis ont été écrits et fonctionnent sous RStudio, version 1.2.1335, utilisant la version 3.6.3 de R. Ils requièrent le chargement des packages suivant :

- `stringr` pour la partie "bases internationales"
- `sqldf` pour les fonctions d'agrégation

2.3 Encodage

Les programmes R ont été encodé en "ISO-8859-1". Cela peut mener à des problèmes de compatibilité. Pour remédier à cela, deux solutions sont possibles. Il est possible d'ouvrir les programmes avec un lecteur de type Notepad et les enregistrer avec un autre encodage. Ou alors, au moment de l'appel des programmes dans le programme principal (utilisation de la fonction "source"), il faut ajouter l'option "`encoding='UTF_8'`".

3 Détail des programmes

Dans cette partie est présenté le déroulé du contenu du programme `tuto_avionic` afin d'aider l'utilisateur à en comprendre les différentes composantes et à pouvoir ainsi utiliser le modèle Avionic pour tous les types de travaux proposés.

3.1 Étapes préliminaires au lancement du modèle

3.1.1 Récupération des sources qui alimentent le modèle

Premièrement, il faut rassembler les données qui vont être mobilisées par Avionic. On peut distinguer ces données de la manière suivante :

- des données de comptabilité nationale : Le Tableau des Entrées Intermédiaires (TEI) importé et domestique, la production, la consommation des ménages importée et domestique, la demande finale importée et domestique, la répartition de l'emploi salarié par produits. Ces données sont disponibles sur le site de l'Insee¹. Un exemple de ces données est disponible sur le Github d'Avionic.
- des données spécifiques à Avionic : les coefficients de transmission et un exemple de variation de demande finale. Elles sont disponibles sur le Github d'Avionic.
- des données nécessaires aux étapes d'agrégation et d'application au compte des ménages par catégories : des tables de passage, une table d'exemple ainsi que la structure du compte des ménages par catégories. Elles sont disponibles sur le Github d'Avionic.
- des données internationales, disponibles sur le site de l'OCDE ou d'Eurostat, par exemple ici : <https://www.oecd.org/sti/ind/inter-country-input-output-tables.htm>

3.1.2 Mise en forme des sources qui alimentent le modèle

Pour être utilisable par les programmes d'Avionic, ces données doivent être mise en forme d'une manière spécifique. Les données issues du Github d'Avionic sont déjà au format requis.

- Les données de comptabilité nationales doivent être placées dans un fichier csv, sans nom de variable. Le séparateur des décimales est un "." et le séparateur des colonnes un ";". Les vecteurs (la production, la consommation des ménages importée et domestique, la demande finale importée et domestique, la répartition de l'emploi salarié par produits) doivent être présentés en ligne.
- Les données spécifiques à Avionic ou nécessaires aux étapes d'agrégation et d'application au compte des ménages par catégories sont déjà au format requis. Les coefficients de transmission sont un vecteur colonne dans un fichier csv. L'exemple de variation de demande finale est un vecteur contenant la variation que l'on veut simuler sur les produits voulus et des 0 pour les autres produits, il contient une seule ligne. Ces deux fichiers contiennent autant de cases que le niveau d'agrégation auquel on travaille². Les tables de passage sont des jeux de coefficients ou des listes d'item en colonnes. La structure du compte des ménages pas catégories est une table par nomenclature fonctionnelle contenant les répartitions des ménages suivant différentes catégories. Il s'agit d'un fichier csv de 32 lignes et 43 colonnes.
- les données internationales sont trouvables sur les sites de l'OCDE ou d'Eurostat. On fournit en exemple des données issues de l'OCDE mise en forme pour être exploitée par les programmes d'Avionic.

3.1.3 Création de l'arborescence, appel des sources et import de données

Une fois au format requis, ces données doivent être regroupées de la manière suivante :

- les données de comptabilité nationale et celles spécifiques à Avionic dans un dossier commun.
- les données internationales dans un dossier commun.
- les données nécessaires aux étapes d'agrégation et d'application au compte des ménages par catégories dans un dossier commun.

De plus, l'ensemble des programmes doivent figurés dans un même dossier.

1. <https://www.insee.fr/fr/statistiques/3547474?sommaire=3547646#titre-bloc-118> pour les données issues du TES symétrique le plus récent, <https://www.insee.fr/fr/statistiques/4131381?sommaire=4131436> pour les données concernant l'emploi.

2. Attention, lorsqu'on travaille au niveau A38 par exemple, il n'y a que 37 colonnes car le produit "loyers imputés" est un sous-produit des activités immobilières

Une fois ces étapes accomplies, on peut commencer à dérouler le programme `tuto_avionic`. Premièrement, on définit les différents chemins des dossiers contenant les données et les programmes. Pour cela on attribue aux variables :

- `chemin_data` le nom du dossier contenant les données de comptabilité nationale et celles spécifiques à Avionic .
- `chemin_data_inter` le nom du dossier contenant les données internationales.
- `chemin_data_agregation` le nom du dossier contenant les données nécessaires aux étapes d'agrégation et d'application au compte des ménages par catégories.
- `chemin_prog` le nom du dossier contenant l'ensemble des programmes.

Dans un premier temps, le programme `tuto_avionic` appelle un ensemble de fonctions de calculs et d'imports. Puis il charge les données de comptabilité nationale et celles spécifiques à Avionic. Pour cela, il faut pour chaque item mettre le nom du fichier correspondant.

3.2 Calcul des contenus de la demande finale en importation, VA et emploi

L'utilisation du modèle proprement dit commence maintenant : les commandes suivantes importent le programme `contenus.R` et l'utilisent pour calculer les contenus en importations, en valeur ajoutée et en emploi de la demande finale.

3.2.1 Contenus en importation de la demande finale

Le programme commence par calculer le contenu en importation de la demande finale. Il s'agit de la somme des contenus directs (consommation finale) et des contenus indirects (consommation intermédiaires de la production consommée). Les contenus indirects sont calculés à partir de matrices de coefficients techniques représentant les quantités de produits nécessaires à la production.

La fonction `contenus_imp` retourne une liste de 6 objets :

- Les contenus en importation en valeur et en part de la consommation des ménages
- Les contenus directs en importation en valeur et en part de la consommation des ménages
- Les contenus indirects en importation en valeur et en part de la consommation des ménages

3.2.2 Contenus en valeur ajoutée de la demande finale

Pour le contenu en valeur ajoutée de la demande finale, la fonction `contenus_va` commence par créer une matrice diagonale de valeur ajoutée pour lui appliquer les coefficients techniques domestiques et la consommation finale domestique. Sont ensuite calculés les contenus en valeur ajoutée directs et indirects. La fonction `contenus_va` retourne une liste de 6 objets :

- Les contenus en valeur ajoutée en valeur et en part de la consommation des ménages
- Les contenus directs en valeur ajoutée en valeur et en part de la consommation des ménages
- Les contenus indirects en valeur ajoutée en valeur et en part de la consommation des ménages

3.2.3 Contenus en emploi de la demande finale

Pour le contenu en emploi (en équivalent temps plein ou en personnes physiques) de la demande finale, la fonction `contenus_emploi` fonctionne de la même manière que `contenus_va`.

La fonction `contenusemploi` retourne une liste de 6 objets :

- Le contenu en emplois total pour chaque produits, ainsi que le nombre d'emploi par million d'euros du produit considéré

- Le contenu direct en emplois pour chaque produits, ainsi que le nombre d'emploi par million d'euros du produit considéré
- Le contenu indirect en emplois pour chaque produits, ainsi que le nombre d'emploi par million d'euros du produit considéré

3.3 Effet d'un choc exogène de demande finale

Le programme `tuto_avionic.R` appelle maintenant le programme `variation_df.R` qui permet de simuler un choc exogène de demande finale en calculant les variations de production, d'importations et d'emplois qui s'ensuivraient. Pour cela, c'est un choc de 5 milliards d'euros sur le premier produit du niveau 38, « Agriculture, sylviculture et pêche » qui est simulé dans la table `var_df` qui est appelée dans les fonctions qui suivent.

Les trois fonctions `variation_prod`, `variation_imp` et `variation_emploi` fonctionnent de la même manière. Elles appliquent une demande finale domestique artificiellement modifiée du choc voulu (le vecteur `var_df`) et calculent les effets directs et indirects sur la production, les importations et les emplois. Ces fonctions retournent :

- La variation de la variable considérée
- La variation directe de la variable considérée
- La variation indirecte de la variable considérée
- La valeur du vecteur calculé en prenant en compte le choc simulé `var_df`

La fonction `variation_imp` retourne aussi la valeur du vecteur des importations sans ce choc de demande finale

3.4 Effet d'un choc exogène de prix

Le programme `tuto_avionic.R` appelle maintenant le programme `variation_prix.R` qui permet de simuler un choc de prix. Cette simulation suppose certaines hypothèses supplémentaires pour être valable³. De plus, les interprétations des résultats seront différentes suivant que l'on utilise un TEI domestique ou bien un TEI total (somme du TEI domestique et du TEI importé). De même, il est possible de réaliser cette simulation en utilisant ou non des coefficients de transmission. Ces coefficients représentent la proportion d'une variation de coût transmise par les entreprises d'une branche à ses clients Voir le document de travail, page 18. La fonction `variation_prix` prend en entrée :

- un TEI (domestique ou total)
- un vecteur de production
- le numéro du produit dont on veut simuler un choc de prix dans la nomenclature. Ainsi pour simuler un choc de prix sur les industries extractives, on utilisera le numéro 2.
- une variation de prix en pourcentage
- `coeff_trans` pour utiliser les coefficients de transmission, rien pour les ignorer

Elle retourne :

- La variation totale, direct et indirecte par produit du prix de ces produits en pourcentage.
- La variation totale du prix de tous les produits en pourcentage.
- L'élasticité prix, soit l'augmentation des prix au total pour une hausse de 1% du prix du produit pour lequel on simule le choc.

3. Voir le document de travail, page 15

3.5 Étude à partir des bases internationales

3.5.1 Découpage international

Pour travailler sur les bases internationales⁴, il faut charger les données internationales. L'exemple sera fait ici sur les bases Tiva (ou ICIO) de l'OCDE, sur 69 pays et 36 produits. Les données peuvent être téléchargées ici :

<https://www.oecd.org/sti/ind/inter-country-input-output-tables.htm>

Le programme permet de charger et répartir les données entre TEI, production, consommation. Il crée aussi la liste des pays, en créant une liste `pays` et une liste `pays2` qui ne contient pas le reste du monde (ROW) et les pays avec données manquantes.

La fonction `decoupage_tei` prend en entrée la matrice des TEI, la liste des pays, le nombre de produits concernés, le nombre de pays et un chemin vers le dossier dans lequel seront créés les fichiers contenant les TEI. Pour chaque pays, sont créés un TEI domestique et un TEI importé depuis chacun des autres pays.

La fonction `decoupage_prod` prend les mêmes éléments en entrée (avec la production au lieu des TEI) et crée les vecteurs de production de chaque pays dans le dossier voulu.

La fonction `decoupage_ef` prend les mêmes éléments en entrée (avec les emplois finals au lieu des TEI) et crée pour chaque pays les emplois finals domestiques et les importés, pays par pays.

Ces découpages sont nécessaires pour des raisons de performance.

3.5.2 Calcul des contenus internationaux

Le programme permet ensuite de calculer les contenus en valeur ajoutée et en importation. La fonction `contenus_va_inter` calcule le contenu en valeur ajoutée d'un emploi final (Consommation des ménages, investissement ou exportations) et le ventile par pays de provenance. On prend ici l'exemple de la consommation des ménages.

Ensuite, la fonction `contenus_imp_inter` calcule le contenu en importation (total, direct et indirect) d'un emploi final et le ventile par pays de provenance. On prend ici l'exemple de la consommation des ménages.

Ces deux fonctions utilisent les tables qui viennent d'être créées par le découpage de la section précédente.

3.5.3 Effet d'un choc exogène de demande finale

Premièrement, on crée une variation exogène de demande finale internationale. On peut créer une variation de demande finale en un ou plusieurs produits, pour un ou plusieurs pays. Si plusieurs produits sont concernés, les variations de production des autres pays sont ventilées selon le pays d'origine de la variation.

Ensuite, la fonction `variation_prod_inter` donne la variation de production (totale, directe et indirecte) qui en découle ainsi que la matrice de production après cette variation de demande finale.

Avec la fonction `variation_imp_inter`, on calcule la variation des importations par pays de provenance et par produit pour le pays qui subit le choc de demande finale simulé, pays que l'on précise comme paramètre de la fonction. Cette fonction fournit aussi les importations par produit et pays de provenance avant et après variation.

4. voir le document de travail, page 34

3.5.4 Effet d'un choc exogène de prix

La fonction `variation_prix_inter` permet de calculer la variation du prix de production des produits suite à un choc exogène de prix pour un produit en paramètre. On simule un choc de prix sur un produit (dont le numéro⁵ est précisé en paramètre) dans un pays dont le numéro dans la liste des pays est précisé en paramètre. Ce choc est exprimé en pourcentage dans les paramètres de la fonction qui retourne les variations (totales, directes et indirectes) par produit et par pays du prix des produits en pourcentage.

3.6 Agrégation et niveaux de nomenclature

La partie suivante décrit des fonctions utiles pour agréger des données à des niveaux de nomenclature moins fins. Par exemple pour passer d'un niveau 38 produits à un niveau 17 produits. Pour cela, il est nécessaire de charger le package `sqldf` qui permet de réaliser des requêtes SQL. Il faut aussi charger les différentes tables de passage qui sont fournies dans le cadre de l'ouverture d'Avionic.

On charge une table `testAgreg` qui va nous permettre de tester les fonctions d'agréations pour une variable exemple `valeur`.

La première fonction `Agreg_NA` permet d'agréger un vecteur à tout niveau de nomenclature NA d'activité économique. Les différents niveaux possibles sont notamment NA10, NA17, NA38, NA64, NA88.

Il faut que le nom de la colonne de nomenclature a agréger porte le nom de sa nomenclature, par exemple "NA17".

La fonction prend en entrée :

- Le vecteur en entrée : il s'agit du nom de la table contenant la variable à agréger ainsi que le niveau de nomenclature de départ
- Le nom de la variable a agréger
- Le niveau de nomenclature initial
- Le niveau de nomenclature voulu (qui doit être moins fin que le niveau de nomenclature initial).

La fonction crée en variable globale une table "`TabAfter`" qui contient les valeurs agrégées pour la variable voulue, renommée "`VAL`"

Ensuite, on teste la fonction `Agreg_NA_fonc` qui permet d'agréger un vecteur de tout niveau NA (produits) en nomenclature fonctionnelle (COICOP⁶ au niveau 1 et 2). C'est particulièrement intéressant pour appliquer le compte des ménages par catégorie et obtenir des décompositions par CSP, âge, ou quintile de revenus.

Cette fonction prend en entrée un niveau de nomenclature de départ et une table comprenant une variable répartie à ce niveau de nomenclature ainsi que le nom de la variable. Dans l'exemple, la variable est une variable "`VAL`" issue de la fonction précédente.

La fonction crée en variable globale une table "`TabAfter_fonc_niv1et2COICOP`" contenant la répartition de la variable suivant la nomenclature fonctionnelle, sous le nom "`VAL_FONC`"

3.7 Compte des ménages par catégories

L'étape précédente est un prérequis à l'étude de la variable considérée sous l'angle du compte des ménages par catégories

5. voir 3.4

6. Classification Of Individual Consumption by purpose ou Classification des fonctions de consommation des ménages

Premièrement, on charge la table de structure du compte des ménages par catégories. Il s'agit d'une table répartissant la consommation des ménages en nomenclature fonctionnelle (COICOP au niveau 1 et 2) suivant plusieurs catégories :

- Quintiles de revenus
- Âges
- Composition du ménage
- CSP
- des croisements de ces variables

Typiquement, on va regarder notre table `"TabAfter_fonc_niv1et2COICOP"` obtenue à l'étape précédente à travers ces catégories.

Pour cela, on utilise la fonction `Ventile_Vecteur` pour répartir notre variable `"Val_fonc"` suivant les différentes décompositions.

On obtient en variable globale une table `Tab_struct_CMpC_foisVecteur` qui est ensuite utilisée en entrée de la fonction `resultat_CMpC`. Cette fonction crée en variable globale une table `TabAgreg` qui contient trois lignes qui correspondent à la décomposition par catégories de la variable étudiée calculée à trois niveaux d'agrégation différent.

La première correspond aux résultats obtenus en agrégeant au niveau 1 de la COICOP, la deuxième en agrégeant au niveau 2, et la troisième en calculant directement à partir de la ligne "Dépense de consommation par ménage" de la table de structure du compte des ménages par catégories, ce qui correspond à une sorte de niveau 0. A priori, les résultats sont plus précis en utilisant l'agrégation issue du niveau le plus fin (ici le niveau 2).

4 Conclusion

Le programme `tuto_avionic` permet donc de mettre en œuvre les principales utilisations d'Avionic dans le cadre de la comptabilité nationale. Notamment les calculs de contenus, les réponses à des chocs de demande finale ou de prix, le tout à un niveau national et international. Il est aussi possible d'étudier les données à travers le compte des ménages par catégories, ainsi qu'à différents niveaux de nomenclature.

Ces programmes sont mis à la disposition de tous, les données sont disponibles sur le site de l'Insee, de l'OCDE (base TiVA) ou d'Eurostat (base FIGARO). Le document de travail expliquant le modèle est aussi disponible sur le site de l'Insee. L'équipe travaillant sur Avionic est joignable à francois.limousin@insee.fr ou alexandre.bourgeois@insee.fr pour des questions ou retours d'expériences.