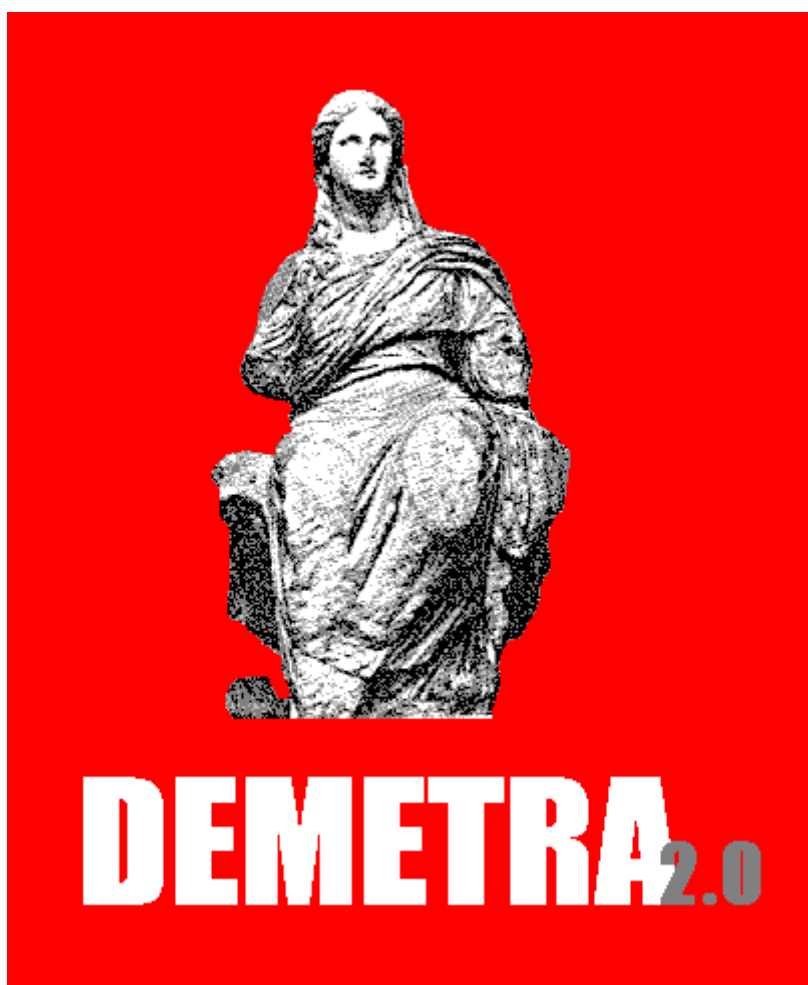


Ajustements saisonniers avec Demetra



Manuel pédagogique

Mai 2002

Eurostat

Office Statistique de la Commission Européenne

SOMMAIRE

I.	LA DÉSAISONNALISATION	3
I.1.	SÉRIE CHRONOLOGIQUE	3
I.1.1.	<i>Stock et flux.....</i>	3
I.2.	DÉCOMPOSITION.....	4
I.2.1.	<i>Saisonnalité.....</i>	5
I.2.2.	<i>Schémas de décomposition</i>	6
I.2.3.	<i>Irrégulier et validation du modèle.....</i>	7
I.2.4.	<i>Événements et valeurs extrêmes.....</i>	10
I.2.5.	<i>Autres effets de calendriers.....</i>	13
I.2.6.	<i>Données corrigées des variations saisonnières.....</i>	15
I.3.	MÉTHODES DE DÉSAISONNALISATION	16
I.3.1.	<i>X12-ARIMA.....</i>	16
I.3.2.	<i>TRAMO-SEATS.....</i>	17
II.	GUIDE D'UTILISATION DE DEMETRA	20
II.1.	CONTEXTES D'UTILISATION.....	20
II.2.	UTILISATION INITIALE.....	22
II.2.1.	<i>Choix de la méthode</i>	24
II.2.2.	<i>Options pour les jours ouvrés</i>	24
II.2.3.	<i>Exploration des résultats du traitement par lot.....</i>	25
II.3.	AMÉLIORATION DES SÉRIES REFUSÉES	29
II.3.1.	<i>Assistant d'amélioration des ajustements refusés</i>	31
II.3.2.	<i>Amélioration des modèles en traitement par lot.....</i>	36
II.3.3.	<i>Système expert.....</i>	44
II.3.4.	<i>Traitement dans le module d'analyse détaillée</i>	45
II.4.	SPÉCIFICITÉS DE L'UTILISATION DE X12-ARIMA.....	48
II.4.1.	<i>Paramétrage de X12-Arima</i>	48
II.4.2.	<i>Exploration des résultats dans le cas de X12-Arima</i>	49
II.5.	STRATÉGIE DE MISE À JOUR ET DE RÉVISION.....	49
II.6.	QUELQUES SITUATIONS TYPIQUES	51
II.6.1.	<i>Changement de saisonnalité sur la période d'étude.....</i>	51
II.6.2.	<i>Non idempotence : saisonnalité dans la tendance.....</i>	55
II.6.3.	<i>Ljung-Box significatif : structure des résidus</i>	57
II.6.4.	<i>Trop d'événements (outliers)</i>	57
II.6.5.	<i>Incorporation d'information a priori</i>	58
II.6.6.	<i>Trois statistiques sur les résidus significatives à 5%.....</i>	65
II.6.7.	<i>Séries non saisonnières.....</i>	66
II.7.	POUR CONCLURE	67



I. LA DESAISONNALISATION

I.1. Série chronologique

Une série chronologique est une suite d'observations d'une variable dans le temps. Dans la quasi totalité des cas, et notamment pour les séries de données sociales ou économiques, l'observation de cette variable se fait à intervalles de temps réguliers et consécutifs : tous les ans, tous les trimestres ou tous les mois. Dans toute la suite de ce manuel, la série chronologique à analyser (« série originale ») sera notée y_t et nous ne traiterons que de données mensuelles ou trimestrielles.

I.1.1. Stock et flux

Il peut s'agir de variables de différentes natures. Pour les besoins de ce manuel, nous distinguerons deux grands types de variables :

- 4 de type stock, ou niveau ; sa valeur est associée à un *instant* donné, par exemple, le solde de votre compte en banque le 30 novembre 2001, la masse monétaire en circulation ou le nombre de demandeurs d'emploi à la même date ;
- 4 de type flux ; il s'agit d'un montant associé à une *période* de temps, ainsi le montant des dépenses que vous avez effectuées dans le courant du mois de novembre 2001, le montant des émissions de billets ou le nombre d'inscriptions au chômage dans le même mois.

Il est important de connaître la nature stock ou flux de la variable. En effet, lorsqu'il est nécessaire de changer d'unité de temps (regroupement des mois en trimestres ou des trimestres en année), les traitements à effectuer sont complètement différents. Ainsi, supposons que nous devions « trimestrialiser » des données mensuelles. Dans le cas d'un stock, nous pourrions retenir soit la valeur correspondant à la valeur observée à la fin du trimestre, soit fournir une moyenne des trois valeurs observées à chaque fin de mois du trimestre. Dans le cas d'un flux, la valeur du trimestre est la somme des valeurs des trois mois.

De même l'unité dans laquelle il faut exprimer la valeur est différente suivant qu'il s'agit d'un stock ou d'un flux. Ainsi, si l'on étudie un montant monétaire, un stock s'exprimera tout simplement en euros (€), ou tout autre unité monétaire, un flux doit s'exprimer en euros par unité de temps (€/mois ou €/an). Dans ce dernier cas, pensez à l'expression de votre rémunération : mensuelle ou annuelle ?

La distinction stock/flux peut être pertinente dans le cas de la désaisonnalisation, car l'impact des effets de calendriers n'est pas le même pour les deux types de variables.

L'étude d'une série chronologique peut être motivée par deux contextes différents :

- 4 la conjoncture,
- 4 la prévision.

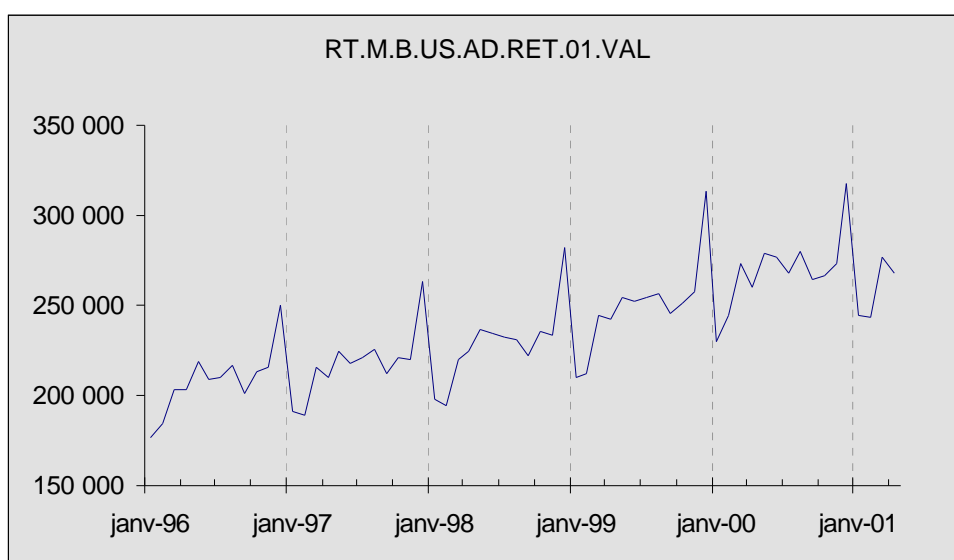
L'analyse conjoncturelle a pour but d'interpréter la situation correspondant aux dernières données observées : que s'est-il passé les derniers mois ? que se passe-t-il aujourd'hui ? est-on en présence d'une croissance, d'une stagnation ou d'une diminution, d'un retournement ? l'évolution observée est-elle structurelle ou accidentelle ?



La prévision a pour but de ... prévoir ! Toutes les méthodes automatiques sont basées sur le postulat implicite que l'information concernant le futur est incluse dans les données du passé et du présent. Elle nécessite donc une analyse de cette information.

I.2. Décomposition

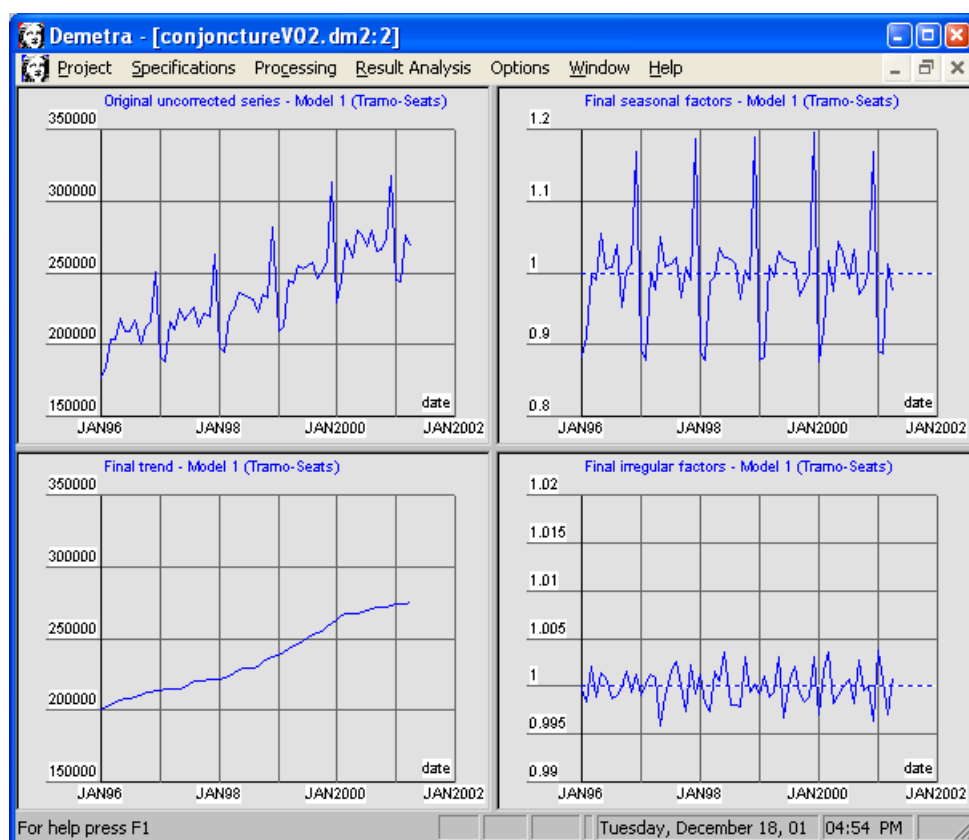
Dans les deux contextes que nous venons d'évoquer, une approche classique consiste à essayer de décomposer la série étudiée en différents éléments ou composantes. Cette approche est très naturelle ; elle est déjà en germe dans la dernière interrogation du conjoncturiste et est tout à fait intuitive lorsque l'on regarde la représentation graphique de séries chronologiques ressemblant à celle-ci (il s'agit, ici, des ventes au détail aux Etats-Unis).



Il est aisé d'en dégager, à l'œil, une décomposition de la série originale (en haut à gauche sur la figure suivante) en :

- 4 une évolution tendancielle (en bas à gauche),
- 4 et une composante saisonnière (en haut à droite).

Le dernier quadrant, en bas à droite, représente le « résidu », c'est à dire ce qui manque pour reconstituer exactement la série originale. Vous noterez que cette reconstitution est pratiquement parfaite, la composante résiduelle laisse apparaître une différence maximale de $\pm 0,4\%$. Attention aux mises à l'échelle automatiques : les deux graphiques de droite représentent des coefficients variant autour de 1, mais celui du haut (composante saisonnière) est à une échelle dix fois plus grande environ que celui du bas (résidus).



La figure a été réalisée en retenant le modèle par défaut de Demetra, c'est à dire en ne faisant que valider les options proposées.

Cette analyse en 3 composantes :

- 4 évolution tendancielle (t_i),
- 4 évolution saisonnière (s_i),
- 4 le reste, noté i_i , pour irrégulier, est à la base des méthodes de désaisonnalisation.

1.2.1. Saisonnalité

Après cette approche expérimentale de la désaisonnalisation, on peut introduire quelques justifications plus théoriques. L'activité humaine est soumise à différents rythmes d'origines dont les effets doivent se retrouver dans le résultat de cette activité. Ainsi, les grands rythmes naturels tels que le rythme circadien (rotation de la terre en 24 heures), la succession des saisons gouvernent des cycles sociaux (alternance du jour et de la nuit, périodes de vacances, etc.) Viennent s'ajouter des rythmes purement sociaux, tels le rythme hebdomadaire et sa conséquence du repos dominical. Le point commun de tous ces cycles est leur caractère parfaitement connu et leur prévisibilité quasi absolue dans le futur. Si l'on dispose de mesures du résultat de cette activité humaine à un rythme infra annuel, il est certain qu'on pourra y retrouver la marque de tous ces rythmes. Il est donc naturel de chercher à en estimer l'impact et d'en tenir compte dans l'analyse des séries chronologiques.

Dans le cas de l'approche conjoncturelle, ceci permettra de dépasser un constat du genre : en été, le chômage diminue et il remonte à l'automne. On va donc *corriger* la série analysée de l'impact des fluctuations saisonnières, on parle alors de données « corrigées des variations saisonnières ». De même, pour la prévision, il sera logique de séparer une tendance, lisse et facilement extrapolable et une composante saisonnière très largement prévisible pour fournir une estimation de l'évolution future



de la série. Ceci suppose, bien sûr, l'hypothèse de la permanence de notre modèle de décomposition dans le futur. Dans ce cas, l'amplitude des variations de la composante résiduelle pourra également être utilisée pour donner une estimation de l'incertitude de prévision et permettre de calculer des « fourchettes » de prévision.

Pour préciser l'analyse, dans le cas de séries moins simples à interpréter que celle que nous venons de considérer, il sera nécessaire d'introduire des composantes supplémentaires.

1.2.2. Schémas de décomposition

Ayant envisagé d'éclater notre série originale en 3 composantes, il faut maintenant nous interroger sur le mode de composition. Diverses solutions sont envisageables. Les plus simples, les plus classiques (et les deux mises en œuvre dans Demetra) sont la composition additive ou multiplicative.

Dans le cas de la composition additive, on supposera que les 3 composantes reconstituent la série originale par la formule suivante :

$$y_t = t_t + s_t + i_t$$

Pour une composition multiplicative, la formule sera :

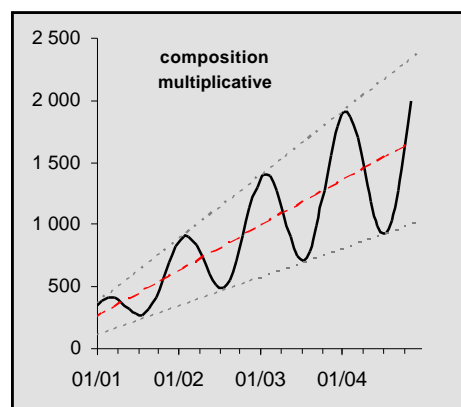
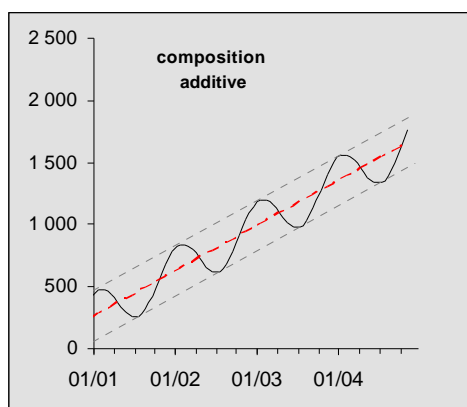
$$y_t = t_t \times s_t \times i_t$$

Pour comprendre la différence entre ces deux mécanismes, on peut considérer les unités des différentes composantes : dans les deux cas, y_t et t_t sont dans les mêmes unités, celles de la variable originale, dans le cas additif, s_t et i_t sont également dans les mêmes unités ; en revanche, dans le cas multiplicatif, elles deviennent de simples coefficients, sans dimension. Ainsi, dans le premier cas, les variations saisonnières sont des variations *absolues*, elle s'exprime sous la forme de + 5 000 € ou – 3 000 €. À l'inverse, dans le cas multiplicatif, les coefficients s'exprimeront sous la forme 1,20 ou 0,95, sans unité, soit, de manière équivalente + 20 % ou – 5 %. Pour faire ressortir ces coefficients en pourcentage, on peut également écrire le modèle de composition multiplicative sous la forme :

$$y_t = t_t \times (1 + s'_t) \times (1 + i'_t)$$

Dans le modèle multiplicatif, les variations saisonnières sont proportionnelles au niveau de la série.

Les graphiques ci-dessous illustrent la différence entre ces deux modèles dans le cas d'une série croissante. Pour un modèle additif, les fluctuations saisonnières sont contenues dans une bande ou un tunnel parallèle à la tendance. Dans le modèle multiplicatif, les fluctuations s'amplifient quand le niveau de la série augmente ; elles font apparaître un cône axé sur la tendance, le cône est divergent quand la série augmente, convergent lorsqu'elle diminue.



C'est en étudiant la représentation graphique de la série qu'il est possible de déterminer le modèle de composition pertinent. On notera que les deux modèles ne sont pas distinguables l'un de l'autre si le niveau de la série reste constant sur l'ensemble de la période. On dit alors que le modèle de décomposition n'est pas identifiable.

Les modèles additif et multiplicatif sont disponibles dans Demetra. Dans Tramo, le modèle multiplicatif est réalisé par la transformation préalable de la série originale par les logarithmes (*log-transform*). L'option par défaut consiste à rechercher la nécessité de cette transformation préalable (*pretest*) ; le modèle retenu est donc, *a priori* additif, sauf s'il est jugé nécessaire de passer à un modèle multiplicatif. Dans la suite de ce manuel, nous supposons, sauf mention contraire que la composition est réalisée de manière additive.

D'une manière générale, il est très important de noter que le problème de l'identifiabilité du modèle de décomposition est un vrai problème : on part d'une série de données et on la décompose en trois séries de données. Il est bien évident qu'il existe une infinité de manières de réaliser une telle décomposition. Elles doivent toutes vérifier l'équation comptable :

$$y_t = t_t + s_t + i_t$$

Pour limiter le champ des possibles et aboutir à une solution raisonnable, il est nécessaire d'explicitier les contraintes qui doivent peser sur les différentes composantes.

La tendance doit être une série plutôt lisse, évoluant sur le long terme et sans accident.

La saisonnalité devra être périodique ou quasi périodique, c'est à dire se reproduisant d'une année sur l'autre de la même manière ou à peu près. On peut envisager d'observer une modification lente de la composante saisonnière si la période étudiée est suffisamment longue. L'effet moyen de la saisonnalité sur une année doit être nul ou, si l'on préfère, le total annuel des valeurs de la série originale doit coïncider avec le total des valeurs de la série hors variations saisonnières sur la même période.

I.2.3. Irrégulier et validation du modèle

La composante irrégulière apparaît alors comme un terme d'ajustement, un résidu, indispensable pour préserver l'équation comptable. Elle joue, cependant un rôle fondamental dans la démarche du statisticien : c'est une clé de la validation du modèle. En effet, ce terme d'ajustement ne doit pas contenir d'information puisque cela invaliderait totalement l'interprétation des autres composantes. En terme d'ingénieurs, on peut s'imaginer qu'il s'agit d'extraire de la série originale un signal, interprétable et de négliger le bruit qui accompagne l'observation. Si après la décomposition, il reste du signal dans le bruit c'est que le modèle retenu est inadapté. En termes statistiques, on considérera



que la composante irrégulière résultante de l'application du modèle doit présenter toutes les caractéristiques de variations aléatoires. On parle parfois de bruit blanc.

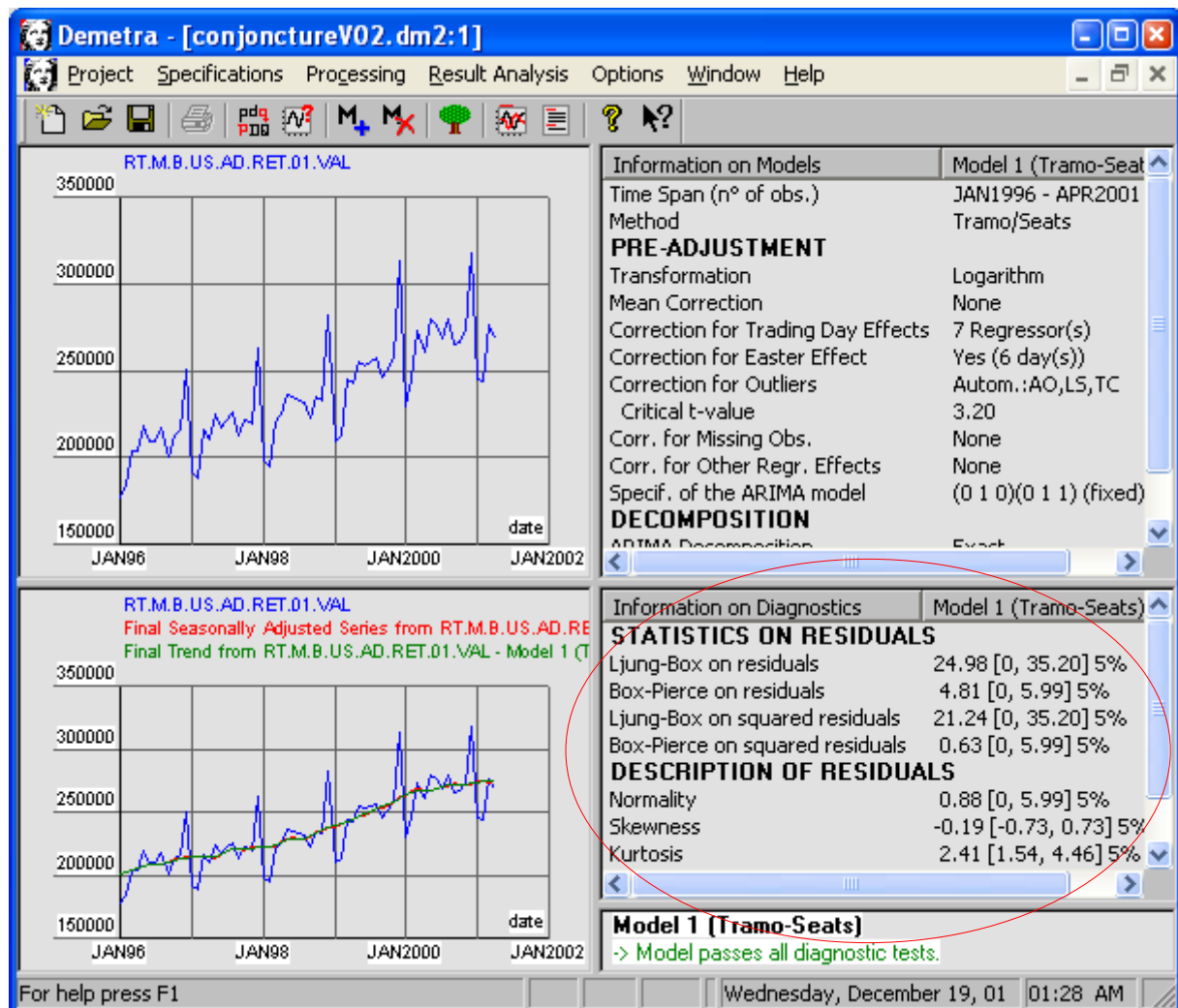
En bref, l'irrégulier, ou composante résiduelle, doit absolument vérifier les hypothèses suivantes :

1. être en moyenne nulle ; en pratique cette hypothèse n'est guère contraignante et est automatiquement vérifiée par les méthodes utilisées.
2. les différentes valeurs de l'irrégulier doivent être indépendantes entre elles. Ceci signifie, entre autres, qu'on ne doit pas être capable de prévoir une valeur du résidu en fonction des valeurs précédentes.
3. être distribuée selon une loi normale. La loi normale ou loi de Gauss est le modèle d'une variable purement aléatoire. C'est la modélisation mathématique d'un bruit sans aucune information.

La vérification des deux dernières hypothèses est impérative pour la validation du modèle de décomposition. C'est pour cette raison que Demetra fait toujours figurer les éléments qui permettent apprécier ces hypothèses dans les résultats fournis pour diagnostiquer le modèle retenu (voir l'image suivante : le coin en bas à droite). Les informations sont présentées légèrement différemment dans le module de traitement automatique et dans le module d'analyse détaillé (version reproduite ci-dessous).

Les valeurs calculées doivent être comprises dans l'intervalle qui suit la valeur ; si ce n'est pas le cas, la ligne correspondante est présentée en rouge.

Les statistiques sur les résidus portent sur le premier jeu d'hypothèses. Une valeur hors de l'intervalle spécifié est un indice qu'il subsiste de l'information dans le bruit. On dit alors que la valeur est significative. La partie description des résidus compare la distribution des valeurs à la distribution d'une loi normale. De la même manière, si l'un des nombres est à l'extérieur de l'intervalle, on peut conclure que la distribution des résidus est significativement différente d'une loi normale.



Enfin, il est utile de comparer l'amplitude des variations saisonnières à l'amplitude des variations dues à la composante irrégulière. En effet, si l'effet saisonnier est noyé dans le bruit, il est probable que l'utilisation d'un modèle de désaisonnalisation pour l'analyse de la série n'est pas indispensable, on dira alors que la composante saisonnière n'est pas significative.

Nous venons de présenter un modèle de désaisonnalisation à trois composantes (tendance, saisonnalité, irrégulier) ainsi que le principe de la démarche de validation d'un modèle, d'où il ressort que l'irrégulier doit impérativement présenter de « bonnes propriétés ». Or, dans cet irrégulier, il peut rester des informations exploitables que notre premier modèle, simple, n'a pas exploité. Il est donc nécessaire de détailler le contenu de notre premier résidu en introduisant de nouvelles composantes.

Il s'agira :

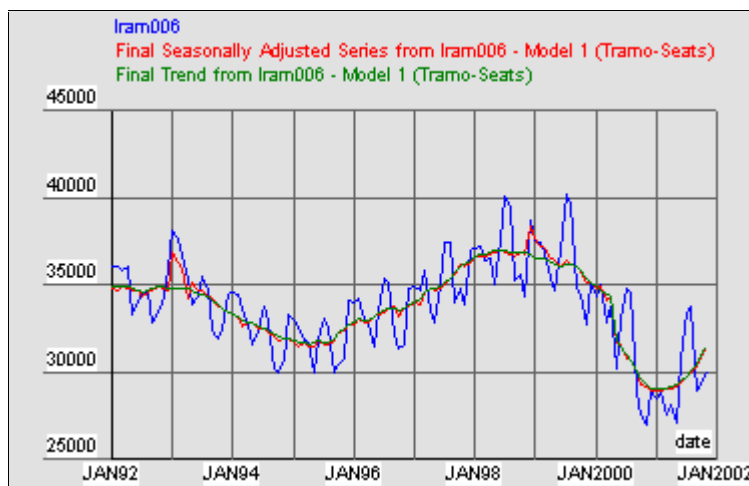
- 4 d'événements ayant marqué l'historique de la série. De nombreux incidents peuvent perturber la série chronologique : pannes, grèves, modification de nomenclature ou de mode de calcul, mesure de relance, effet d'annonce. Ces incidents peuvent être connus *a priori* et introduits dans le modèle par le statisticien ; ils peuvent aussi être détectés automatiquement par Demetra. Il peut alors, être intéressant d'essayer d'identifier les événements qui sont repérés par Demetra.
- 4 d'autres effets de calendrier ne se ramenant pas à une variation saisonnière : correction pour jours ouvrés, effet de la date de Pâques ou correction pour les années bissextiles. Les caractéristiques du calendrier sont connues et prévisibles, leurs effets peuvent avoir une incidence sur le niveau de la série, Demetra peut les détecter et les estimer.



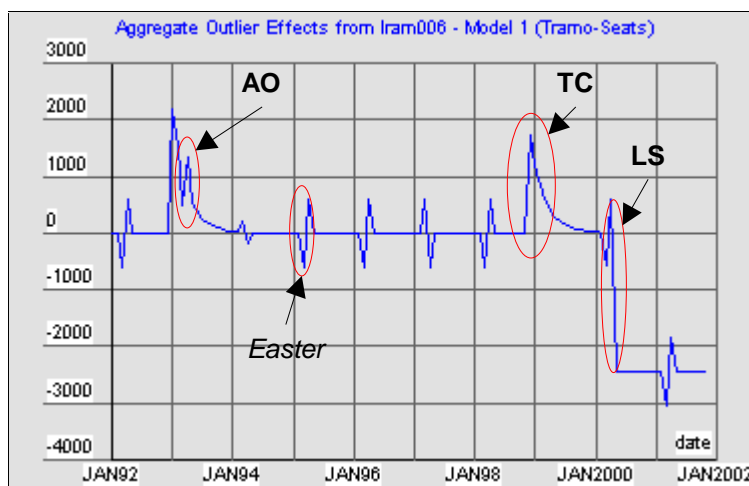
Ces différentes composantes vont fournir autant de corrections préalables. Elles sont résumées dans la volet en haut à droite de la fenêtre de projet sous la rubrique *Pre-Adjustment*. Pour la série présentée dans la copie d'écran précédente, la plupart de ces options ont été activées.

1.2.4. Événements et valeurs extrêmes

Dans l'exemple suivant (nombre de femmes bénéficiaires d'allocation de chômage en Irlande, les options par défaut de Demetra, modèle Tramo-Seats, donnent les résultats suivants : désaisonnalisation (acceptée) et tendance :



Terme d'ajustement *a priori* :



Ce dernier graphique est obtenu par l'option « *Aggregate Outlier Effects* ». Il regroupe les différents événements détectés automatiquement, AO, LS et TC ainsi qu'un effet de calendrier (*Easter*). On peut noter que certains d'entre eux sont suffisamment notables pour être repérés sur la représentation de la série originale, de la série corrigée des variations saisonnières et de la tendance.

On aboutit ainsi à une décomposition de l'irrégulier en deux sous composantes :

- ev_t : regroupant les « événements », c'est-à-dire l'ensemble des corrections *a priori*,
- i_t^* : le « vrai » résidu, ne comprenant plus qu'une composante aléatoire.



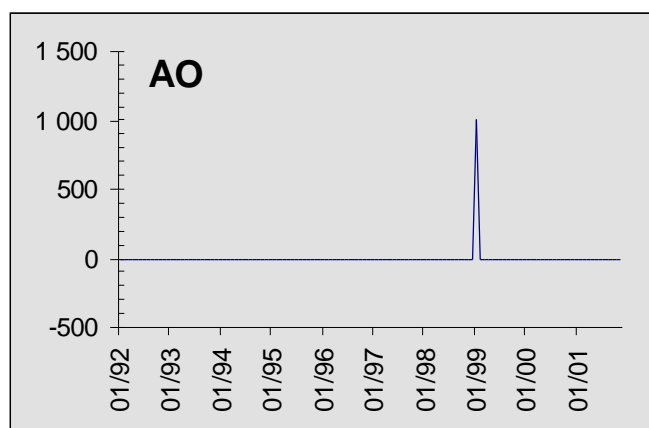
Les événements décrits par le terme de correction *a priori* sont, en principe, identifiables, soit « susceptibles d'être identifiés ». Dans l'idéal, le conjoncturiste devrait les identifier ; soit qu'il y retrouve des événements bien connus de la vie économique (grève importante, intempéries violentes), soit qu'il parte à la recherche de ce qui a pu causer un impact sur la série étudiée à la date repérée. Dans ce dernier cas, on pense souvent, en premier lieu, à de « vrais » événements, au sens d'événements climatiques, sociaux ou politiques. Mais il ne faut pas négliger les « faux » événements : les artefacts liés au processus de production des données étudiées. Il peut s'agir ici d'après peu près tout ce que l'on peut imaginer : du changement de nomenclature à l'erreur de saisie ou de transmission, ou, pour rester dans l'environnement de l'exemple présenté, de la destruction de sacs postaux dans une agence de l'emploi lors d'une inondation à la mise en œuvre sur une durée déterminée d'une politique de radiations plus actives.

Comme on peut le voir aux quelques exemples cités, l'identification effective des accidents, c'est-à-dire leur interprétation demande une bonne connaissance de l'environnement socio-économique mais aussi des règles et conventions de définition des données étudiées, de leur évolution et de l'environnement administratif dans lequel les données sont produites.

Dans le cadre de ce manuel, et, malheureusement, dans celui de la pratique courante de l'analyste ou du conjoncturiste, on se limitera à décrire les différents types d'accident susceptibles d'apparaître. Chacun est présenté ci-dessous, en version additive, à la même date et d'une valeur caractéristique identique.

Ø Valeurs extrêmes (AO)

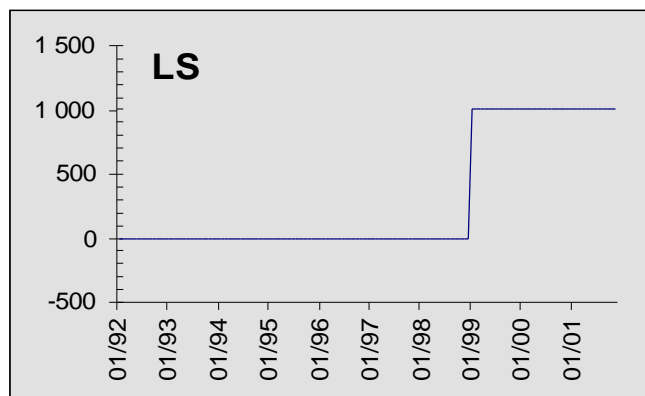
Une valeur extrême, en anglais *additive outlier*, correspond à un point isolé. Il peut s'agir de l'effet d'une grève, d'un événement météorologique ou inondation, etc. Dans le modèle de décomposition, ce type d'événement est intégré à l'irrégulier.





Ø **Changement de niveau (LS)**

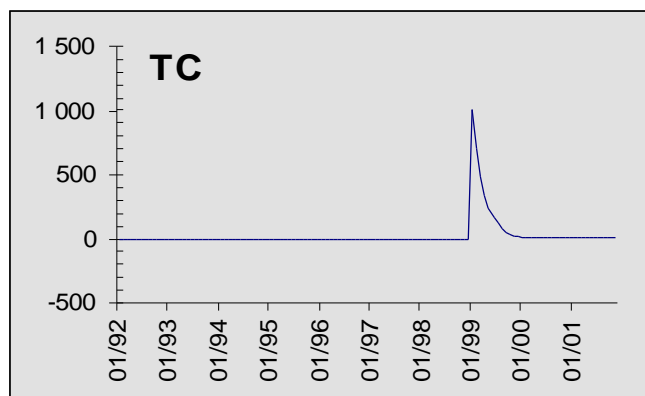
Le changement de niveau (*level shift*) ou échelon correspond à une rupture dans la valeur moyenne de la série. Le changement reste acquis. Ce type d'événement survient, par exemple, lors de changement de définition de la série, modification de la nomenclature...



Le changement de niveau est traité par Demetra comme une composante venant modifier la tendance.

Ø **Choc transitoire (TC)**

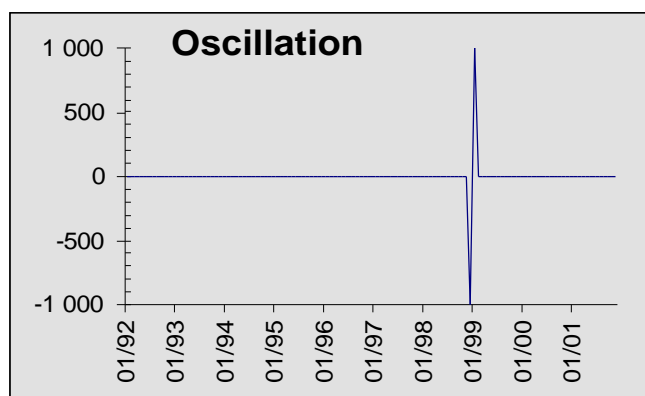
Le choc transitoire décrit une perturbation avec un retour amorti à la situation de départ. Pour le décrire, il faut comme pour les précédents une taille, mais aussi une durée d'amortissement.



Le choc transitoire est un événement de type « irrégulier ».



On peut remarquer qu'il manque un accident que l'on peut trouver fréquemment dans certains types de données : l'oscillation.



Elle apparaît par des phénomènes de report d'une période sur la suivante. Elle peut résulter d'effet d'annonce : mise en place d'une prime à compter d'un mois donné ou, dans le sens inverse, date de fin d'attribution de la prime.

Bien que n'étant pas considéré comme un type d'événement, on peut retrouver (ou décrire) une oscillation sous la forme de valeurs extrêmes consécutives en sens inverse. Notons d'ailleurs que l'effet Pâques se repère aisément grâce à une telle oscillation : il s'agit tout simplement d'un report d'une période sur la suivante lorsque la date de Pâques tombe à la fin du mois de mars (de mars sur avril, en données mensuelles ; du premier au deuxième trimestre, en données trimestrielles).

Ainsi, pour conclure cette partie, on peut considérer que notre décomposition s'écrit maintenant sous la forme :

$$y_t = t_t + s_t + ev_t + i^*_t$$

dans la version additive, avec une formule équivalente pour un modèle multiplicatif.

I.2.5. Autres effets de calendriers

Plusieurs raffinements sont possibles pour la prise en compte de la saisonnalité et, plus généralement, pour intégrer les effets de calendrier.

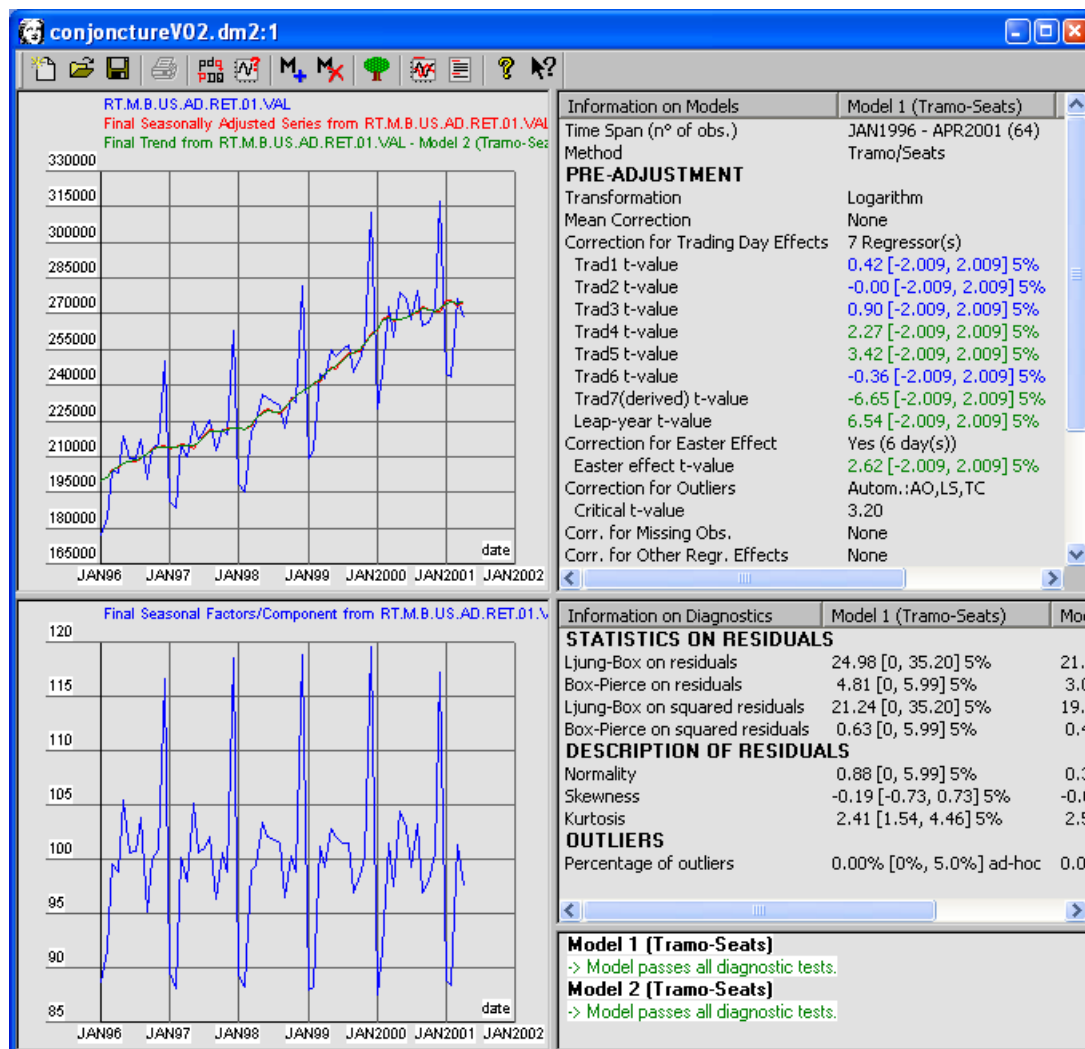
Les différents mois du calendrier, comme les quatre trimestres de l'année, sont de longueurs différentes (28, 29, 30 ou 31 jours pour les mois ; 90, 91 ou 92 jours pour les trimestres). Ceci peut entraîner des conséquences pour les données de type flux, dont la valeur dépend de la longueur de la période. Habituellement, cette influence est incorporée dans l'effet saisonnier « pur ». Mais il est également possible, dans X12, de distinguer l'effet saisonnier pur, lié au mois, de l'effet longueur du mois. Enfin, les mois ont, en général, la même longueur d'une année sur l'autre, *sauf* le mois de février, il peut donc être nécessaire d'introduire une correction pour les années bissextiles (*leap year*).

Les mois, par ailleurs, ne sont pas composés d'un nombre entier de semaines. À l'exception des mois de février non bissextiles, ils comprennent donc un, deux ou trois jours supplémentaires. Les différents jours de la semaine ne figurent donc pas tous le même nombre de fois dans chaque mois : certains y sont 4 fois, d'autres 5 fois. Or, une grande partie de l'activité sociale est conditionnée par le rythme hebdomadaire : traditionnellement, on se marie le samedi, le repos dominical est la règle en Europe, en France, beaucoup de commerces sont également fermés le lundi, etc. On voit donc que la composition du mois en jours de semaine peut avoir une incidence directe sur le niveau de la série.



Ici aussi, cette remarque est particulièrement fondée pour les données de type flux. À titre d'exemple, le niveau des ventes dans la grande distribution en décembre est directement lié au nombre de samedis dans le mois ; le mois de décembre comptant 31 jours, il y a donc, *a priori*, 3 années sur 7 comportant 5 samedis et 4 n'en comportant que 4. La comparaison de ces mois de décembre entre eux doit impérativement tenir compte de ce paramètre. Ces corrections sont regroupées sous le terme générique de « correction pour jours ouvrés » ou composantes (*trading day effects*). Elles aboutissent à calculer un coefficient pour chacun des jours de semaine décrivant l'intensité de l'activité pour ce jour de semaine relativement à la valeur moyenne.

Dans l'exemple ci-dessous, présentant l'évolution mensuelle des Ventes au détail aux États-unis, on peut constater la présence d'une correction pour jours ouvrés : ainsi, le dimanche (jour 7) présente un facteur très négatif, les jeudi et vendredi sont les jours de plus fort achat. La somme des effets des sept jours de la semaine est, par construction, nulle.

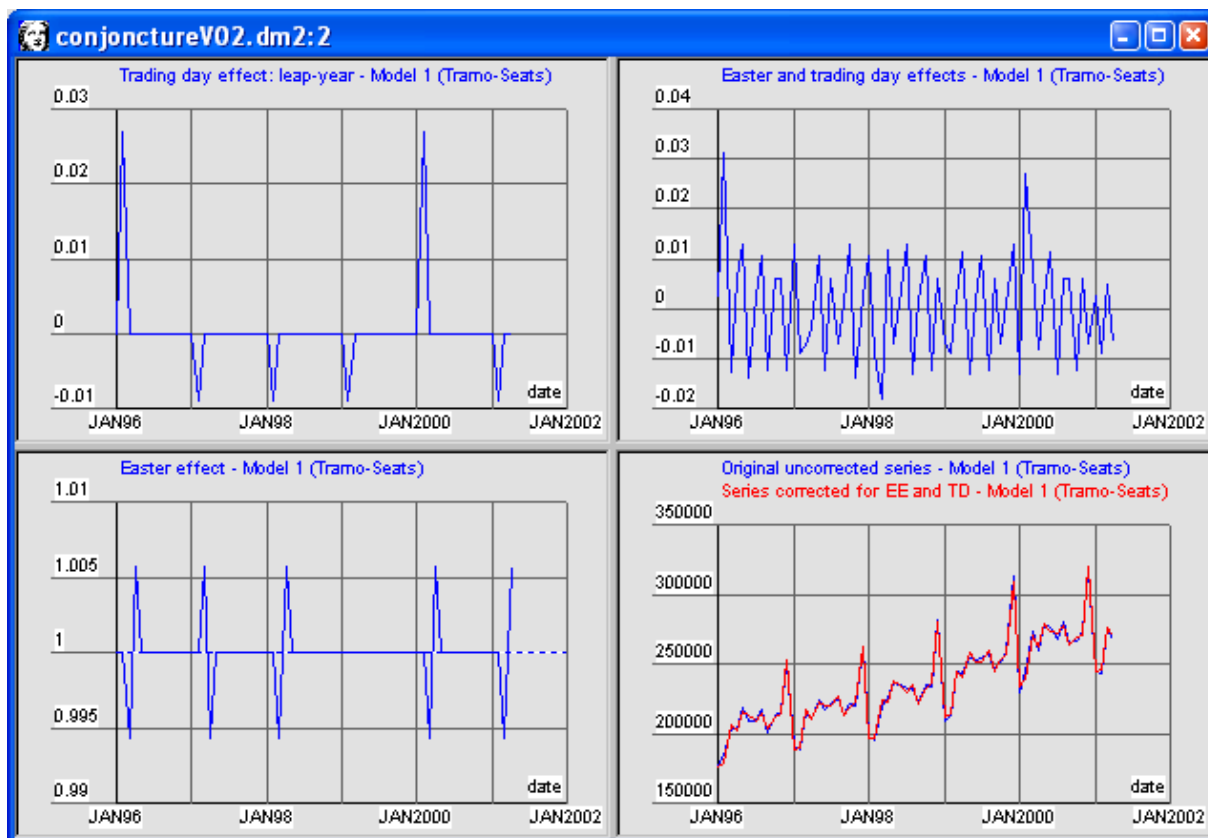


On note également la présence d'un effet Pâques et d'un effet année bissextile. La copie d'écran ci-dessous, extraite du module d'analyse détaillée, présente les différentes composantes liées aux effets de calendrier :

- 4 en haut, à gauche, l'effet de l'année bissextile, tous les mois de février, nul en moyenne sur 4 ans, positif les années bissextiles, négatif les trois autres.
- 4 en bas, à gauche, l'effet Pâques,



- 4 en haut, à droite, l'ensemble des corrections liées au calendrier, autres que les variations saisonnières. Elles regroupent les 2 composantes précédentes plus l'ensemble des effets des poids de chaque jour de la semaine.
- 4 enfin, le dernier graphique permet de comparer la série originale et la série corrigée des effets de calendrier. Lorsque ces corrections sont pertinentes, la série corrigée devrait avoir une allure plus régulière que l'originale : la périodicité devrait être plus marquée.



Notons pour terminer que Demetra propose, par défaut, de tester la nécessité de ces différentes corrections et de ne les appliquer que si les effets estimés sont jugés significatifs. Dans l'exemple ci-dessus, les effets sont jugés significatifs si leur valeur est située à l'extérieur de l'intervalle mentionné (pour tous les paramètres, l'intervalle vaut $[-2,009 ; +2,009]$). Plus la valeur s'en éloigne, plus l'effet est important. En ce qui concerne les poids des jours de semaine, il suffit que l'un des jours ait un effet significatif pour que la correction des jours ouvrés soit retenue.

1.2.6. Données corrigées des variations saisonnières

De façon traditionnelle, on appelle série corrigée des variations saisonnières, abrégé en cvs, en anglais *seasonally adjusted*, la série initiale dont on a retiré l'impact de la composante saisonnière. Cette donnée, bien sûr non observée, permet donc de disposer d'une estimation de ce que serait la valeur de la série compte non tenu de l'effet saisonnier.



En appliquant cette définition avec les notations de notre décomposition et en faisant fonctionner l'équation comptable, on obtient les différentes versions suivantes :

$$\begin{aligned}cvs_t &= y_t - s_t \\ &= t_t + i_t\end{aligned}$$

On remarquera donc que la cvs s'interprète comme la résultante de la tendance ET de l'irrégulier. Autrement dit, elle comprend la tendance, mais aussi les fluctuations, aussi bien aléatoires que correspondant à des événements. La pratique, aujourd'hui bien passée dans les mœurs, de la publication de données cvs, fait qu'il est courant de considérer que cette donnée est confondue avec une indication de tendance. Lorsque le conjoncturiste publie la donnée cvs, il doit être conscient qu'elle nécessite, encore, une interprétation : est-il possible de discerner, pour cette dernière valeur observée, la part de ce qui est structurel de ce qui est conjoncturel ? La réponse est, évidemment, non ! et il serait illusoire d'espérer trouver la réponse dans la consultation des (nombreux) paramètres et indicateurs sortis par le modèle : si un événement nouveau est en train de se produire actuellement, ou depuis quelques mois, on ne voit pas comment le modèle, aussi sophistiqué soit-il, pourrait en trouver son origine dans le passé.

Un exemple récent éclairera cette position. Si le conjoncturiste constate que ses données concernant le mois de septembre 2001 sont « atypiques », c'est à lui de s'interroger sur la conjoncture et de chercher autour de lui, et en relation avec la série étudiée, si un événement ne se serait pas produit qui pourrait avoir une influence sur ce qui est décrit par cette série. Rien, dans aucun modèle, ne lui dira ce qu'il doit chercher. C'est sa connaissance de la situation économique, politique, sociale, météorologique ainsi que sa connaissance du fonctionnement du secteur étudié qui lui suggérera les hypothèses qu'il lui est possible d'émettre.

I.3. Méthodes de désaisonnalisation

Demetra permet de choisir, éventuellement pour chaque série étudiée, entre deux méthodes appartenant à des familles bien distinctes :

- 4 les méthodes basées sur une modélisation économétrique de la série chronologique, ou méthodes paramétriques. C'est le choix SEATS de TRAMO/SEATS ; la méthode a été proposée en 1996.
- 4 des méthodes d'origine empirique, ne nécessitant pas le recours explicite à un modèle, dites aussi méthodes non paramétriques. Elles ont une origine ancienne, basées essentiellement sur des méthodes de type moyenne mobile. Demetra propose la méthode X12-ARIMA, dérivée, en 1996 également, d'une lignée qui passe par X11, en 1965 puis X11-ARIMA en 1975.

I.3.1. X12-ARIMA

Dans son principe, cette méthode est basée sur la technique des moyennes mobiles. Nous allons la présenter rapidement en supposant des données mensuelles.

La composante saisonnière étant supposée périodique ou quasi-périodique, d'effet annuel moyen nul, l'idée vient très naturellement de calculer les moyennes de la série sur une période de 12 mois consécutifs. Le résultat de ce calcul, rattaché à la période de temps médiane (le mois situé au milieu de la période des 12 mois) est donc « filtré » de sa composante saisonnière, on dispose donc d'une approximation de la tendance.

En retirant cette tendance de la série brute, on calcule des écarts à la tendance. L'équation comptable de décomposition nous indique alors que ces écarts à la tendance comprennent une composante



saisonniers et un terme résiduel. L'étape suivante consiste alors à regrouper les écarts correspondant à une même saison (un même mois) et à en calculer des moyennes mobiles (par exemple, la moyenne des écarts du mois de mars pour 5 années consécutives). Les valeurs ainsi obtenues peuvent alors être utilisées comme estimation des coefficients saisonniers.

À partir de ce principe plutôt simple, les méthodes de types X11 ou X12 vont empiler la répétition de cette démarche un certain nombre de fois. Ce sont les « étapes » : après une étape A, préliminaire, les étapes B, C et D vont répéter chacune deux fois ces techniques de filtrage.

Une partie de la complexité de cet empilement provient du fait que dans les étapes B et C, la méthode utilisée cherche à détecter des points extrêmes (AO) en essayant de repérer les valeurs trop importantes de la composante irrégulière. La série est alors « corrigée » de ces valeurs extrêmes et passée en entrée de l'étape suivante.

Un autre point délicat des méthodes de type X11 se présente lorsqu'il s'agit de s'intéresser aux extrémités de la série étudiée. Si le point de départ n'est pas forcément l'enjeu de débats importants, la dernière valeur observée, en revanche revêt une importance particulière : y compris dans une approche conjoncturiste, c'est bien pour les derniers points que se posent avec le plus d'acuité le problème d'interprétation de la décomposition en tendance-saison-irrégulier. Or, comme nous l'avons indiqué plus haut, la moyenne mobile sur 12 mois induit nécessairement un décalage de 6 mois dans l'estimation de la tendance.

Deux solutions sont proposées à cette difficulté des fins de séries :

- 4 utiliser une moyenne mobile différente de la simple moyenne arithmétique équipondérée (la moyenne des 12 valeurs). Ceci implique d'introduire une pondération de chacune des 12 valeurs dont on doit calculer la moyenne et cette pondération aboutit à augmenter l'influence des tout derniers points de la série. Or ces points sont *a priori* moins bien connus : ils peuvent être provisoires ou soumis à révision. Cette méthode est la méthode initialement retenue dans X11.
- 4 l'autre solution consiste à prolonger la série de données aux deux extrémités. L'idée est d'ajuster un modèle aux données et de s'en servir pour extrapoler (ou rétropolier, vers le passé) ces données afin de disposer de points (fictifs) dépassant les extrémités observées et sur lesquels on peut appuyer un calcul de moyenne mobile non (ou moins) dissymétrique. La prévision repose sur les méthodes de Box et Jenkins, méthodes d'identification et d'estimation de modèles ARIMA. C'est ce qui explique le nom de X11-ARIMA ou X12-ARIMA.

Bien que l'on retrouve ici une technique de modélisation à la base de TRAMO/SEATS, il est important de comprendre que son intervention, dans le cadre des méthodes non paramétriques, se fait à un niveau beaucoup moins fondamental : il s'agit plutôt d'une méthode pour améliorer l'estimation des extrémités de série. De plus, il est tout à fait envisageable que X12-ARIMA n'arrive pas à ajuster un modèle ARIMA pour compléter ses points ; ceci ne perturbe pas la méthode qui se rabat alors sur l'algorithme initial.

I.3.2. TRAMO-SEATS

À l'inverse des méthodes « empiriques » présentées ci-dessus, SEATS repose fondamentalement sur la modélisation de la série de données. La famille des modèles utilisables est fournie par la théorie des processus stochastiques et est connue sous le nom de modèles ARIMA. Dans de tels modèles, la valeur de la série à un instant donné est déterminée par plusieurs composantes :

- 4 une partie dite « auto-régressive », c'est le AR de ARIMA, ce qui signifie que la valeur est déterminée par une relation avec les valeurs précédentes de la série. Le nombre de valeurs précédentes (ou retard, *lag* en anglais) utilisées dans cette relation est dit l'ordre de la partie AR. Il est souvent noté par la lettre p , dans les modèles utilisés, il peut varier de 0 à 3.
- 4 une partie dite « moyenne mobile », *moving average* ou MA. Il s'agit d'une composante aléatoire. Elles fait intervenir la valeur d'une variable de bruit, mais cette variable peut être un bruit sans

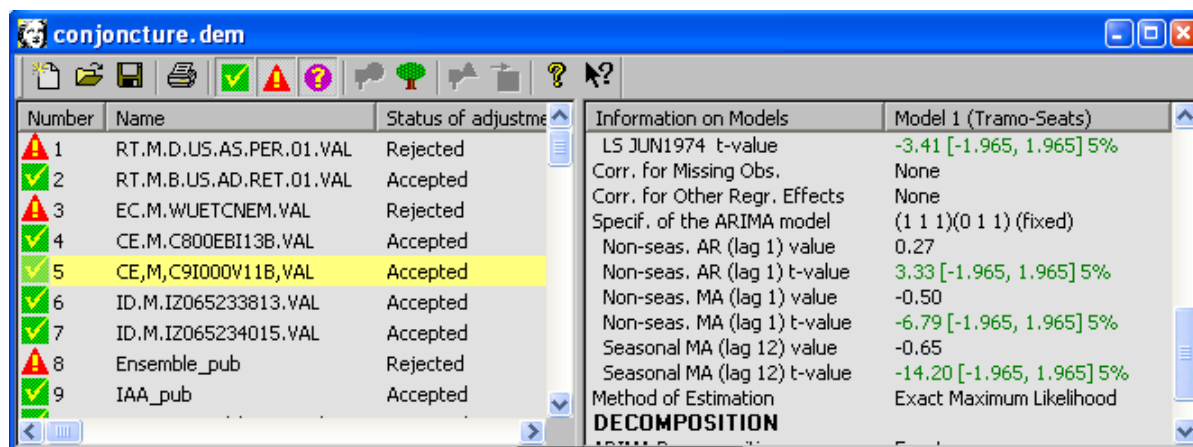


structure, ou bruit blanc, ce qui signifie que les différentes valeurs du bruit sont indépendantes entre elles, ou, au contraire, posséder une certaine structure : les termes aléatoires successifs sont corrélés entre eux. L'étendue temporelle de cette structure est décrite par l'ordre de la partie MA, noté q . Il indique le nombre de retards à utiliser pour déterminer la partie aléatoire à partir d'un bruit blanc. Habituellement, sa valeur est comprise entre 0 et 3.

- 4 enfin, le I , pour *integrated*, indique si l'on doit utiliser la série elle-même ou ses variations (ou dérivées). L'ordre de différenciation, noté d , indique si le modèle utilise simplement la variable elle-même ($d = 0$), s'il doit inclure sa dérivée (ses variations, $d = 1$) ou sa dérivée seconde ($d = 2$, valeur maximale utilisée par SEATS).

L'intégration de variations saisonnières dans les modèles ARIMA se fait en greffant sur le modèle de base présenté ci-dessus un autre modèle dit saisonnier. Le modèle de base traite de retards exprimés en nombre de valeurs observées, c'est-à-dire en mois pour des données mensuelles ou en trimestres pour des données trimestrielles. Le modèle saisonnier décrit le lien entre la valeur de la série à un moment donné et sa valeur l'année précédente. Les retards sont maintenant exprimés en nombre d'années. Pour décrire cette partie du modèle, on introduit donc trois nouveaux paramètres correspondants aux ordres respectifs des parties auto-régressive, intégrée et moyenne mobile du modèle saisonnier. Ils sont notés respectivement sp , sd , sq et sont limités aux valeurs 0 et 1 dans SEATS.

En résumé, le type de modèle utilisé pour décrire la série étudiée est décrit (on dit spécifié) par la donnée de deux triplets $(p, d, q)(sp, sd, sq)$. Cette information sur le modèle utilisé est indiquée dans les sorties de Demetra. La phase de sélection d'un modèle est dite l'identification du modèle. Par défaut, Demetra se charge d'identifier le modèle à utiliser pour décrire la série, mais laisse la possibilité à ceux qui maîtrisent la théorie d'intervenir dans le choix du modèle. Dans un deuxième temps, les méthodes ARIMA doivent estimer les coefficients de ce modèle. Le nombre de coefficients à calculer correspond exactement aux ordres retenus, hors différenciations.



Number	Name	Status of adjustment	Information on Models	Model 1 (Tramo-Seats)
1	RT.M.D.US.AS.PER.01.VAL	Rejected	LS JUN1974 t-value	-3.41 [-1.965, 1.965] 5%
2	RT.M.B.US.AD.RET.01.VAL	Accepted	Corr. for Missing Obs.	None
3	EC.M.WUETCNEM.VAL	Rejected	Corr. for Other Regr. Effects	None
4	CE.M.C800EBI13B.VAL	Accepted	Specif. of the ARIMA model	(1 1 1)(0 1 1) (fixed)
5	CE.M.C9I000V11B.VAL	Accepted	Non-seas. AR (lag 1) value	0.27
6	ID.M.IZ065233813.VAL	Accepted	Non-seas. AR (lag 1) t-value	3.33 [-1.965, 1.965] 5%
7	ID.M.IZ065234015.VAL	Accepted	Non-seas. MA (lag 1) value	-0.50
8	Ensemble_pub	Rejected	Non-seas. MA (lag 1) t-value	-6.79 [-1.965, 1.965] 5%
9	IAA_pub	Accepted	Seasonal MA (lag 12) value	-0.65
			Seasonal MA (lag 12) t-value	-14.20 [-1.965, 1.965] 5%
			Method of Estimation	Exact Maximum Likelihood

Dans la figure ci-dessus, la série sélectionnée est décrite par un modèle d'ordre $(1, 1, 1)(0, 1, 1)$. Nous trouvons donc trois coefficients : AR(1) et MA(1), de valeur 0,27 et -0,50, associés aux ordres p et q $(1, 1, 1)(0, 1, 1)$ et MA(12) correspondant à l'ordre sq $(1, 1, 1)(0, 1, 1)$, de valeur -0,65. Leurs valeurs sont suivies des indicateurs de significativité (t -value) ; dans le cas d'un modèle retenu automatiquement par Demetra, elle sont significatives (t -value à l'extérieur de l'intervalle) pour les coefficients correspondants à l'ordre le plus élevé de chaque partie.

Une des difficultés des méthodes de type ARIMA découle de la sensibilité de la méthode aux perturbations. En effet, si une série chronologique subit des événements ou des ruptures (voir ci-dessus, AO, LS, TC), ceux-ci vont influencer fortement sur les calculs nécessaires à l'identification du modèle et l'estimation des paramètres. Il est donc très important de mettre en place une détection fine de telles perturbations afin de garantir la qualité du modèle retenu. C'est le rôle de la partie TRAMO de TRAMO/SEATS. C'est également la raison pour laquelle les corrections *a priori* de la série sont



plus détaillées dans l'option TRAMO/SEATS que dans la version de X12-ARIMA implémentée dans Demetra.



II. GUIDE D'UTILISATION DE DEMETRA

II.1. Contextes d'utilisation

Demetra a été conçu par des statisticiens, dans un environnement de conjoncturistes, à destination de conjoncturistes. L'organisation du logiciel est donc fondée sur une hypothèse d'organisation du travail du conjoncturiste. Celui-ci reçoit les dernières mises à jour des données brutes et doit en fournir une interprétation ; cette tâche récurrente est susceptible de concerner un nombre important, voire très important, de séries. Pour cela, il utilise, entre autres, des méthodes de désaisonnalisation. Il lui faut donc :

- 4 choisir une méthode et un modèle,
- 4 appliquer ce modèle pour calculer les données cvs.

La différence entre ces deux aspects est perceptible dans l'écart de sens des traductions française et anglaise : *seasonal analysis* ou désaisonnalisation.

Ces deux volets sont soumis à des cycles distincts : l'utilisation du modèle, la désaisonnalisation, se fait mensuellement pour la publication des chiffres, le choix du modèle ou la réestimation de ses paramètres ne se fait pas, en général, à chaque publication mais à un rythme moins fréquent, typiquement de l'ordre de une fois par an.

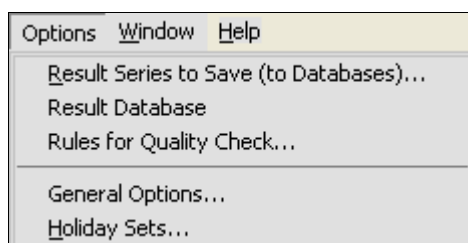
Le volume important des données à traiter justifie le recours à une automatisation maximale, tant pour l'analyse que pour la désaisonnalisation. C'est le rôle du module de traitement par lot (*Automated module*). Il vise à couvrir les besoins courants de production et suppose une organisation des séries en « projet »

Le module d'analyse détaillée (*Detailed analysis module*) permet de réaliser une étude particulière sur une série pour laquelle le statisticien dispose de temps à consacrer à la recherche d'un modèle adapté ; soit que la série soit particulièrement importante, soit qu'elle présente des particularités fortes nécessitant une intervention plus poussée. Il permet de tester différentes méthodes ou différents choix de paramètres sur une même série, de conserver les résultats de ces expériences et d'en réaliser une comparaison notamment sous forme de graphiques.

Nous allons suivre la démarche de production et accompagner le conjoncturiste dans son travail. Nous commençons par la définition d'un nouveau « projet », celle-ci nécessite :

- 4 la définition des séries à traiter : choix du fichier contenant les données et la détermination de la période à analyser
- 4 le choix des résultats à stocker,
- 4 ainsi que celui de l'emplacement (fichier) où les stocker,
- 4 les statistiques utilisées et les règles de sélection des modèles.

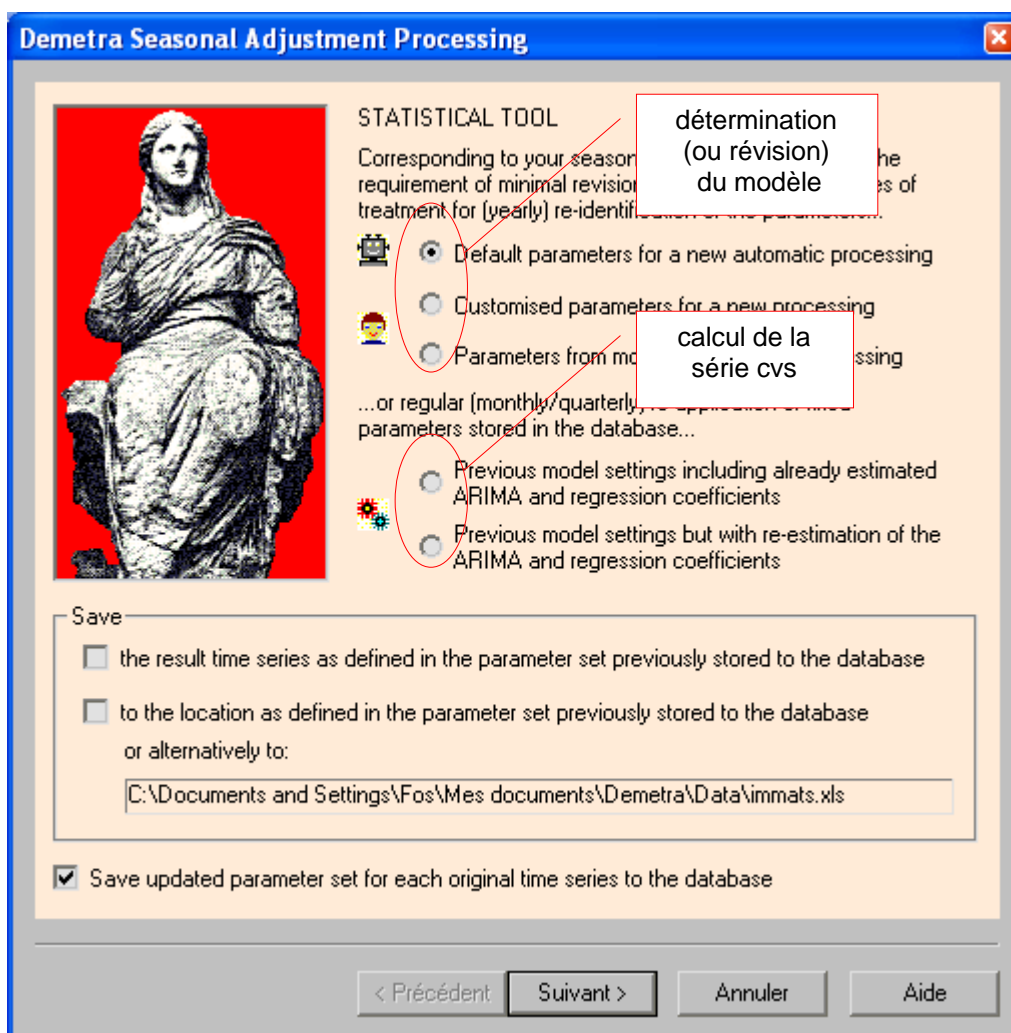
Ces choix sont réalisés au cours des cinq écrans de l'assistant de création de projet (*Demetra Project Wizard*) le dernier point faisant appel à deux écrans. Tous ces choix sont communs aux séries d'un même projet et stockés dans la définition du projet. Ils peuvent être modifiés ultérieurement en passant par le menu **Options**. Les trois premiers éléments de ce menu correspondent exactement aux quatre écrans de l'assistant nouveau projet.



Les différentes options générales peuvent être modifiées ultérieurement à l'aide du menu ci-dessus. La période d'analyse est modifiable chaque fois que l'on relance un nouvel ajustement.

Après avoir paramétré le projet, vient le moment de l'analyse (dialogue *Seasonal Adjustment Processing*). Il faut d'abord choisir le type de traitement à appliquer :

- 4 *seasonal analysis* (identification et estimation du modèle) ou
- 4 désaisonnalisation (calcul des données corrigées des variations saisonnières).



En effet, ces deux niveaux d'activité sont bien distincts dans le dialogue initial.

Les trois premiers boutons radio de choix correspondent à la détermination des modèles de décomposition saisonnière (*seasonal analysis*). Le terme *new processing* indique qu'il s'agit de



l'identification, ou ré-identification, du modèle de désaisonnalisation. Les trois options proposées correspondent :

- 4 au choix par défaut. En gros, Demetra teste la nécessité des différentes corrections possibles et les inclut s'il estime qu'elles améliorent le modèle ;
- 4 le deuxième choix permet à l'utilisateur de fixer lui-même les différentes corrections à mettre en œuvre ou à désactiver ;
- 4 le troisième suppose que ces choix ont été préalablement effectués et stockés dans un fichier décrivant le modèle.

Les deux options suivantes correspondent à la phase « désaisonnalisation » :

- 4 application du modèle aux nouvelles données (quatrième bouton),
- 4 calcul des données corrigées des variations saisonnières après recalcul (réestimation) des coefficients du modèle.

Les cases à cocher correspondent à des options de stockage des résultats des calculs (les deux premières) et des paramètres des modèles (la troisième). Vous remarquerez que leur position par défaut dépend du choix de traitement effectué par les boutons radio. Lors d'une estimation des modèles, c'est la troisième case qui est cochée ; pour un calcul de désaisonnalisation, ce sont les deux premières. Il y a peu de raisons de modifier ce choix préétabli.

II.2. Utilisation initiale


Continuons à accompagner l'utilisateur dans sa démarche. Nous supposons que c'est la première fois qu'il traite le groupe de séries du projet. Il a donc choisi un traitement impliquant la recherche du modèle de décomposition. Il va maintenant devoir déterminer la méthode à utiliser : Tramo-Seats ou X12-Arima, puis en fonction de ces choix précédents, les différents paramètres d'utilisation de ces méthodes. Le tableau ci-dessous résume les possibilités d'intervention en fonction du type de traitement choisi.

	traitement choisi			
	automatique		manuel	
options de l'utilisateur	Tramo/Seats	X12-Arima	Tramo/Seats	X12-Arima
effets de calendrier (cjo)	x	x	x	x
jours fériés	x		x	
transformation de la série			x	x
paramétrage X11				x
valeurs extrêmes			x	x
correction biais cvs			x	x
choix modèle ARIMA			x	x

Toutes les options listées correspondent à un dialogue (ou onglet d'un dialogue) à l'exception de la fixation des jours fériés qui est une des options du dialogue « effet de calendrier ». Vous aurez noté que la version actuelle de Demetra exclut la possibilité de fixer la liste des jours fériés pour X12-Arima.



Plaçons-nous dans l'optique d'un traitement le plus automatisé possible (presse-bouton !), il nous reste donc deux choix : la méthode et la correction des jours fériés.



NEW AUTOMATIC SEASONAL ADJUSTMENT

Seasonal Adjustment Method

☒ Tramo/Seats
☐ X-12-Arima

Modelling Time Interval

...

 /

 to

...

 /

☒ Otherwise reload from the saved parameters, if available

Type of Trading Day Effect to Test

☐ No trading day adjustment
0 regressors

☐ Working days (Monday to Friday):
1 regressor

☐ Working day (Monday to Friday) & leap-year:
2 regressors

☐ Trading day (Monday, Tuesday, ..., Saturday):
6 regressors

☒ Trading day (Mon, Tue, ..., Sat) & leap-year:
7 regressors

Country-Specific Holidays

☒ Allow reducing the number of trading day regressors


< Précédent

Suivant >

Annuler

Aide





NEW AUTOMATIC SEASONAL ADJUSTMENT

Seasonal Adjustment Method

☒ Tramo/Seats
☐ X-12-Arima

Type of Trading Day Effect to Test

<input type="radio"/> No trading day adjustment	0 regressors
<input type="radio"/> Working days (Monday to Friday):	1 regressor
<input type="radio"/> Working day (Monday to Friday) & leap-year:	2 regressors
<input type="radio"/> Trading day (Monday, Tuesday, ..., Saturday):	6 regressors
<input checked="" type="radio"/> Trading day (Mon, Tue, ..., Sat) & leap-year:	7 regressors

Country-Specific Holidays

< Précédent

Suivant >

Annuler

Aide

II.2.1. Choix de la méthode

La question de la méthode reste ouverte : Eurostat a mené de nombreuses comparaisons entre les deux méthodes et a conclu qu'il n'était pas possible de conclure à la supériorité de l'une des deux méthodes. Le choix dépend donc essentiellement des choix ou habitudes de l'utilisateur et de l'organisme pour lequel il travaille. D'un point de vue purement pragmatique, on peut noter que sur quelques essais comportant de nombreuses séries, où seul le traitement par défaut est utilisé (aucune intervention de l'utilisateur) Tramo-Seats accepte un peu plus de séries que X12-Arima ; d'autre part, des séries rejetées par l'une des méthodes peuvent être acceptées par l'autre. Enfin, pratiquement, la possibilité d'introduire les jours fériés dans le modèle est une facilité supplémentaire de Tramo-Seats que, malheureusement, Demetra ne propose pas pour X12-Arima.

Pour terminer, remarquons que le choix de la méthode comprend un aspect « idéologique » : les débats entre modélisateurs (les modèles ARIMA de Tramo-Seats) ou tenants de l'approche empirique (X11 ou X12) pouvant être animés...

II.2.2. Options pour les jours ouvrés

Les options proposées pour la correction pour les jours ouvrés sont les suivantes :

- 4 pas de correction,



- 4 jours ouvrés (1 régresseur). Cette option revient à ne considérer que deux types de jours : les jours ouvrés, du lundi au vendredi et les jours non ouvrés, samedi et dimanche. Elle revient à compter d'un côté les jours ouvrés du mois et de l'autre les jours de week-end. Cette option correspondrait à une option correction du nombre de jours ouvrés au sens strict du terme.
- 4 estimation d'un poids pour chaque jour de la semaine (6 régresseurs). C'est l'option la plus souple qui permet de pondérer chaque jour. Elle paraît donc devoir être retenue *a priori*, cependant, il ne faudrait pas la retenir si l'on pense qu'il n'y a pas de différence sensible entre les différents jours ouvrés de la semaine, car en surcorrigeant, elle réintroduirait, des fluctuations aléatoires.

La correction étant activée, il est possible et, sans doute, souhaitable d'incorporer également la correction pour l'année bissextile (choix à 2 ou 7 régresseurs). Cette variante est à conseiller.

II.2.3. Exploration des résultats du traitement par lot

Les options ayant été déterminées, il ne reste plus qu'à lancer le traitement. Le résultat s'affiche alors pour l'ensemble des séries dans la partie en haut à gauche de la fenêtre ; le statut du modèle s'affiche sous forme d'icône au début de la ligne et il est possible, à l'aide de la barre d'outils, de filtrer l'affichage pour ne retenir que les séries acceptées, rejetées ou restant à traiter. Si le temps à consacrer à l'étude des séries est limité, on ne retiendra que les séries rejetées, en faisant confiance aux options par défaut de Demetra.

Une toute première remarque : en utilisant Tramo-Seats, certaines séries sont rejetées, avec un message d'erreur :

Error returned from X12-Arima. Skipped! Error in the input parameters.

Ce message est produit lors de la phase de recherche d'un modèle ARIMA. En explorant les séries comme indiqué ci-dessous, il est probable que l'on découvre des séries qui ne devraient pas figurer dans le projet, ainsi dans différents exemples, nous avons rencontré des séries dont la valeur est constante et égale à 0. X12-Arima ne sait pas traiter de telles séries. Pour d'autres séries correspondant à de véritables données à traiter, il serait souhaitable d'employer la méthode X12 sans la partie ARIMA. Malheureusement, ceci n'est pas possible ; nous suggérons donc d'essayer d'utiliser la méthode X12-Arima sur ces séries en empêchant l'utilisation des prévisions (forçage à 0 de *Number of Forecasts* dans le dialogue de paramétrage personnalisé).




Demetra Customised Seasonal Adjustment Processing

Outliers & External Regressors Bias Correction & Trend Smoothing

Log Transformation Calendar Effect Corrections

Arima Model, Mean Correction & Forecast



ARIMA model

- ☒ Automatic identification and estimation
- ☐ Automatic selection and estimation
- ☐ Estimation of a specific ARIMA model

Mean correction

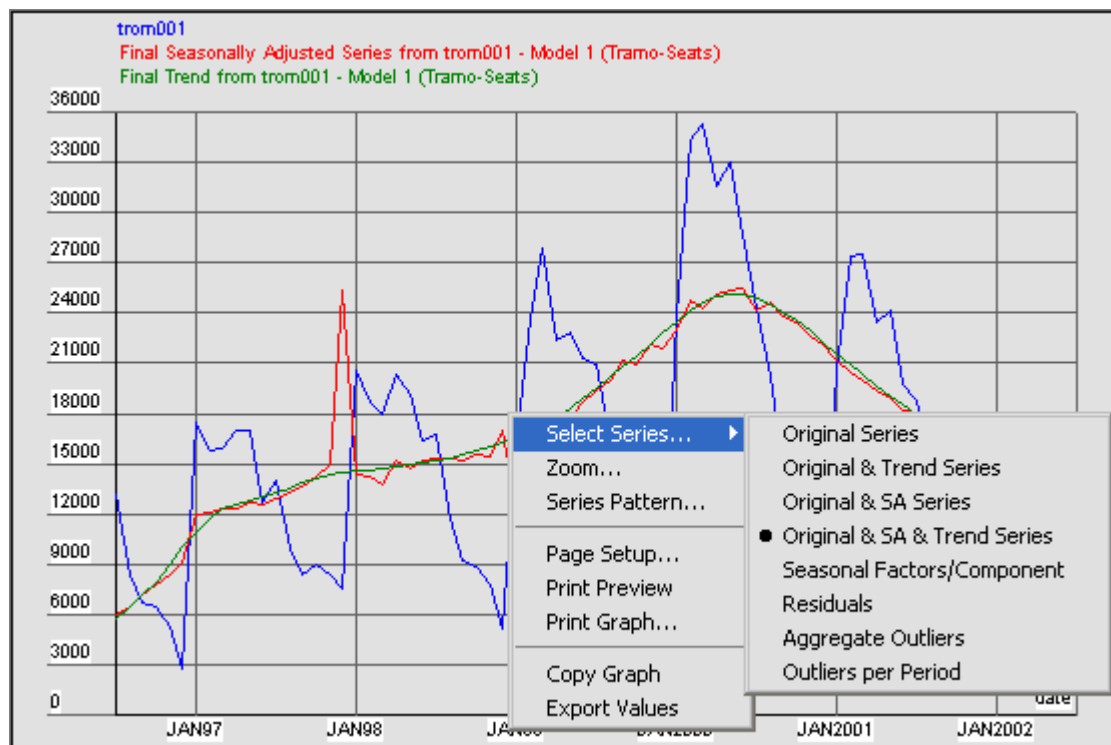
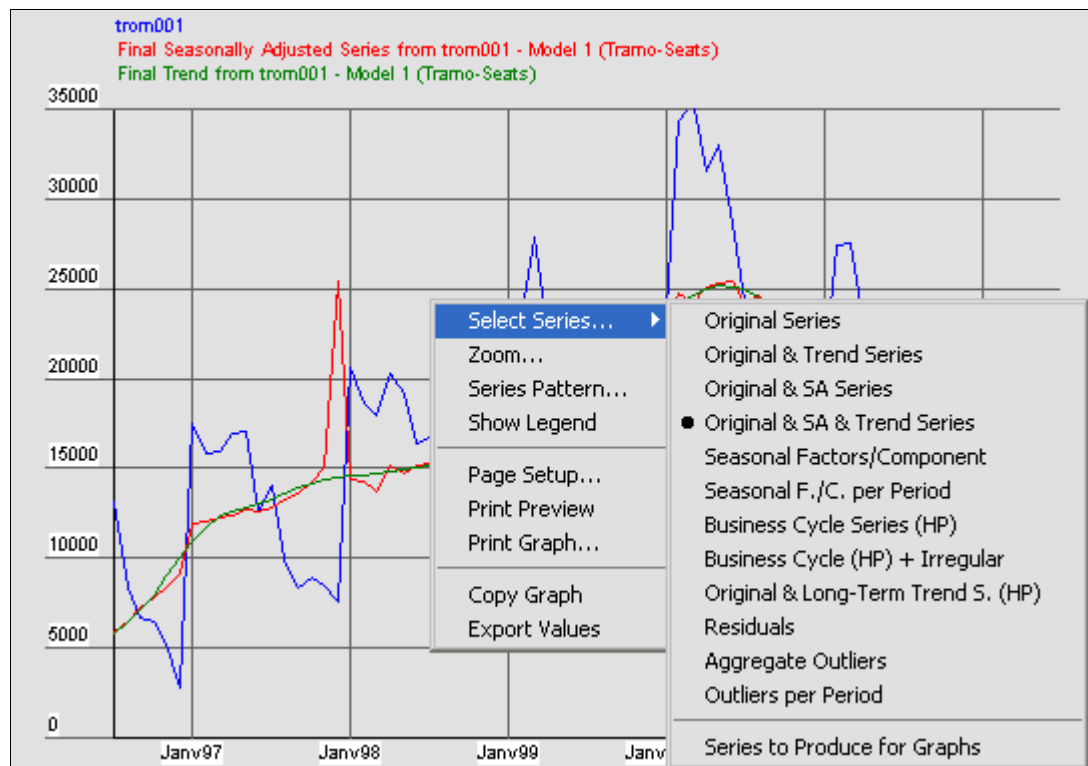
- ☒ Yes
- ☐ No

Number of Forecasts 0

OK Annuler Appliquer Aide

Ø Exploration de la série brute

Si l'on souhaite avoir une vision d'ensemble des résultats et, parfois, entrevoir les difficultés rencontrées au cours de l'estimation du modèle, il est utile de parcourir l'ensemble des séries en visualisant les principaux graphiques dans la partie située en bas à gauche. Pour ce faire, il suffit de cliquer avec le bouton droit dans cette partie et de sélectionner la ou les courbes à représenter.

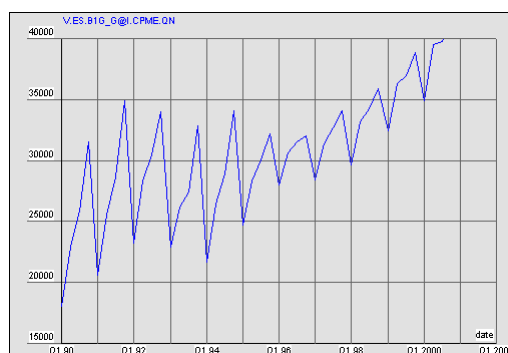
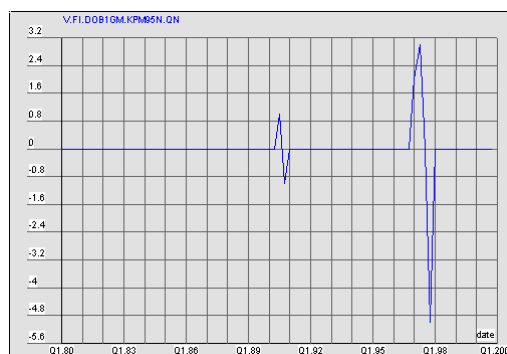
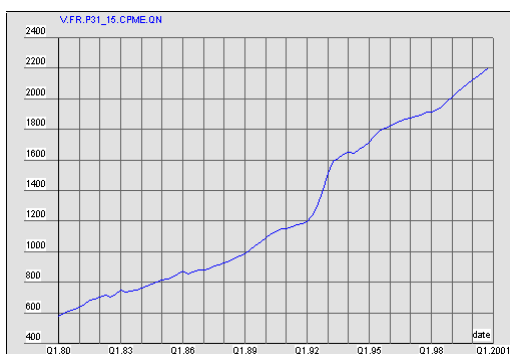


Puis l'on revient dans la liste des séries en cliquant sur la première. On peut alors faire défiler les séries en utilisant les flèches de déplacement vers le haut ou le bas.

L'un des premiers choix consiste à retenir l'ensemble des trois séries originale, cvs et tendance. Par exemple, dans la figure reproduite ci-dessus, l'écart important et ponctuel entre la série corrigée des variations saisonnières et la tendance permet de repérer instantanément la présence d'une valeur extrême en décembre 1997.



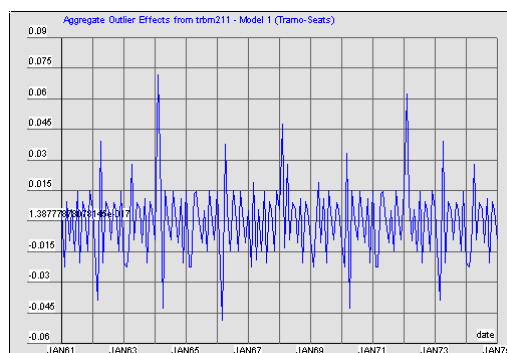
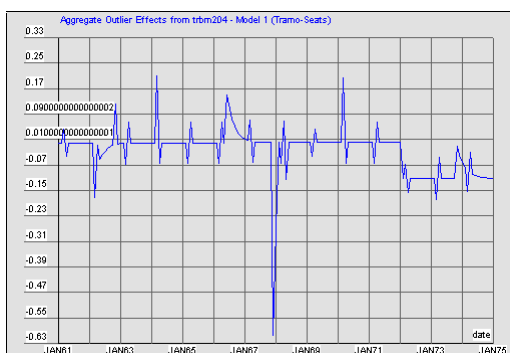
Le parcours des représentations graphiques de la série brute peut être l'occasion de découvrir des informations qui auraient pu échapper à la sagacité du conjoncturiste : ayant reçu plusieurs centaines de séries, il n'a pas nécessairement eu le loisir de les regarder. Le défilement, même rapide, des graphiques, autorisé par l'interface de Demetra, risque de faire apparaître des séries telles celles qui sont représentées ci-dessous (elles sont toutes issues d'un traitement réel).



Le premier cas de figure correspond à une série ne présentant aucun caractère saisonnier, le deuxième est une série très particulière correspondant, vraisemblablement, à des ajustements ou report d'une période sur la suivante. Le troisième graphique traduit la présence d'un changement de mode de fonctionnement : au premier trimestre 1995, il y a eu changement brusque de la composante saisonnière ; dans ce cas, il n'est pas aisé de tenir compte de cette rupture et il est recommandé de scinder la série en deux, et de désaisonnaliser séparément les deux morceaux de la série. En pratique, cela signifie qu'il faut supprimer du projet les points de la série correspondants à la première partie.

Exploration des événements

De la même manière, l'étude (rapide) de la composante saisonnière, du résidu et des termes de correction a priori (*Aggregate outliers*) est susceptible d'apporter des informations utiles sur les séries étudiées. Voici, ci-dessous, quelques exemples de sorties graphiques des corrections a priori.



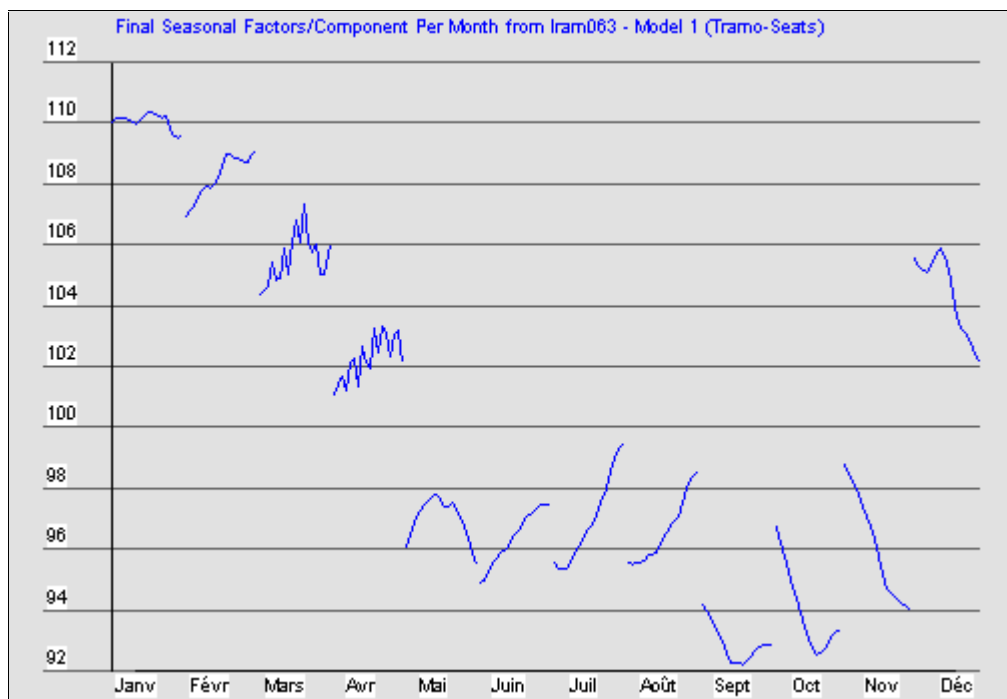


À gauche, on repère quelques événements (AO, TC, LS) et, annuellement, la répétition d'une petite oscillation correspondant à l'effet Pâques. À droite, l'aspect presque aléatoire traduit la présence d'un effet jours ouvrés : il y a donc un terme correcteur pour chacun des mois, correspondant à sa composition en fonction des différents jours de semaine. On peut distinguer la présence de points extrêmes (AO), mais il est plus délicat de repérer qu'un effet Pâques est également présent. L'aspect de ce graphique est, bien sûr, fonction de l'importance relative des différents éléments de correction *a priori*. Notons, pour terminer que le graphique peut ne pas apparaître (message : **The selected series is not available**), ceci signifiant simplement qu'aucun événement n'a été détecté. La version actuelle de Demetra ne permet pas de le représenter, alors qu'existent d'autres effets de calendrier (correction pour jours ouvrés, par exemple).

Dans l'idéal, en supposant que le temps du conjoncturiste ne soit pas limité, il faudrait identifier les différents événements, c'est-à-dire retrouver les causes socio-économiques ayant provoqué les variations dans les valeurs de la série.

Ø Exploration de la saisonnalité

L'option *Seasonal F/C per period* permet de suivre l'évolution sur l'ensemble de la période étudiée des coefficients saisonniers.



On peut ainsi juger de la déformation de la structure saisonnière sur longue période. Ainsi, dans l'exemple ci-dessus (demandeurs d'emploi en Irlande de 1980 à 1997), on peut constater que le creux saisonnier de l'été (juin, juillet, août) a tendance à se combler alors que le creux de l'automne s'est encore accru.

II.3. Amélioration des séries refusées

Que faire des séries refusées lors de notre premier traitement ? Et tout d'abord, éliminons d'emblée, la solution, expéditive, qui consisterait à modifier les critères d'acceptation (et donc de rejet) d'un modèle. Ceci peut se réaliser, nous l'avons vu, lors de la définition du nouveau projet, mais peut être modifié, *a posteriori*, dans le menu **Options/Rules for quality check...**



Demetra Project Wizard - Rules for Quality Check of X-12/Arima Adjustments
Demetra Project Wizard - Rules for Quality Check of Tramo/Seats Adjustments

Statistics checked for significance

☒ Ljung-Box on residuals ☒ Ljung-Box on squared residuals

☒ Box-Pierce on residuals ☒ Box-Pierce on squared residuals

☒ Normality ☐ Skewness

☐ Kurtosis

A time series adjustment is rejected

☒ if the decomposition of the chosen ARIMA model was not admissible

☒ if there is at least 1 of the selected statistics which is significant at a 0.1 % level

☒ if there are at least 3 of the selected statistics which are significant at a 5 % level

☒ if the automatic outlier detection procedure detected more outliers than 5 % of the number of observations of the original time series

Set to Default Reinitialize

Demetra Project Wizard - Rules for Quality Check of Tramo/Seats Adjustments
Demetra Project Wizard - Rules for Quality Check of X-12/Arima Adjustments

A time series adjustment is rejected

☒ if none of the ARIMA models was chosen

☒ if the Ljung-Box statistic is significant at a 0.1 % level

☒ if the average percentage standard error in within-sample forecasts over the last year is greater than 15 %

☒ if the automatic outlier detection procedure detected more outliers than 5 % of the number of observations of the original time series

☒ if the combined statistic Q (M1, M3-M11) is not accepted

Set to Default Reinitialize

Ce choix, peut-être tentant, n'est pas raisonnable ! Il aboutit à éviter délibérément les difficultés rencontrées, à ignorer l'information indiquant la nature de ces difficultés et, finalement, à faire comme si les modèles étaient satisfaisants, ce qu'ils ne sont pas.

De la même manière, il est, bien sûr, possible de forcer l'acceptation d'une série dont le modèle est refusée. Cette option ne doit être utilisée qu'en dernier recours. Dans la plupart des cas, il est possible de trouver la raison du problème et de conclure quant au traitement qu'il faut appliquer, quitte à ce qu'une des réponses possibles soit : ne pas désaisonnaliser, car l'allure de la série ne le justifie pas !

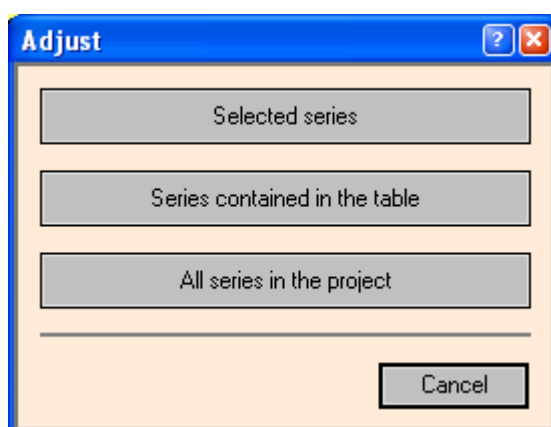
Plus sérieusement, pour améliorer les séries refusées, Demetra fournit dans ses menus deux possibilités :

- 4 relancer un ajustement (menu *Processing/Redo adjustment*)
- 4 ou l'amélioration des séries refusées (menu *Improvement of rejected adjustment/Start improvement of selected series*).

Ces deux options sont présentes dans la barre d'outils de Demetra. Elles sont voisines et sont entourées d'un cercle sur l'image ci-dessous



L'option relancer un ajustement permet de se retrouver face au dialogue initial de paramétrage d'un modèle de désaisonnalisation. Étant donné que nous avons déjà effectué une passe de traitement automatique, il nous faudra, pour cette seconde passe opter pour un ajustement personnalisé (deuxième option du dialogue). Ce nouvel ajustement du modèle saisonnier peut s'appliquer à plusieurs ensembles de séries, en fonction du choix effectué dans le dialogue suivant :



L'ajustement s'appliquera,

- 4 soit aux séries sélectionnées,
- 4 soit à l'ensemble des séries affichées dans la table (en tenant compte donc des filtres éventuellement appliqués à l'affichage)
- 4 soit à l'ensemble des séries du projet (donc affichées ou non dans la table). Cette option correspond plutôt à une réestimation de l'ensemble des séries plutôt qu'à la poursuite de notre estimation initiale.

Nous utiliserons plutôt la première option, mais après quelques manipulations que nous verrons un peu plus loin.

II.3.1. Assistant d'amélioration des ajustements refusés

En effet, Demetra propose une façon de traiter toujours de manière automatisée, mais avec quelques variantes par rapport à l'approche standard initiale des séries qui auraient été rejetées lors du premier traitement. Il s'agit de l'amélioration des séries refusées (*Improvement of rejected adjustments*). Cette option ne vous est accessible que si, parmi les séries sélectionnées dans la table du projet, l'une au moins d'entre elles a été refusée.

Le choix de cette option dans le menu (ou de l'outil correspondant dans la barre d'outils) fait apparaître une nouvelle fenêtre. Celle-ci reprend dans ses cinq volets les informations de la fenêtre de projet à l'exception du volet contenant la liste des séries du projet qui est remplacée par une autre représentation graphique. Il est donc possible de faire figurer plusieurs graphiques complémentaires décrivant la série à traiter ; typiquement, et par défaut, la série brute, la tendance et la série cvs, dans le volet en haut à gauche et la composante saisonnière dans le volet en bas à droite. À la différence de ce qui se passe dans la fenêtre de projet, il est maintenant possible de stocker l'information concernant plusieurs modèles différents, à condition qu'ils soient tous de la même famille que le modèle initial (Tramo-Seats ou X12-Arima) : en effet, chaque fois que vous testez une amélioration, un nouveau modèle est ajouté dans le volet droit de la fenêtre.

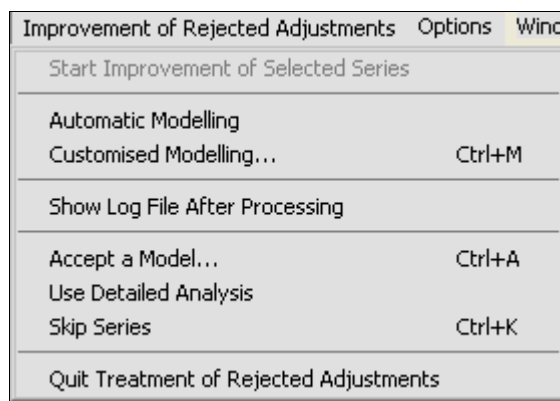
Les séries de votre sélection dans la table du projet dont le modèle a été rejeté vont maintenant vous être proposées, une par une. À vous d'en améliorer la modélisation, le logiciel vous assistant dans la recherche d'un nouveau modèle. Après utilisation de ces possibilités, vous pourrez alors :

- 4 accepter l'un des modèles testés
- 4 basculer dans le module d'analyse détaillée
- 4 passer (*skip*, ignorer) la série.



Attention, le passage à l'analyse détaillée vous fait quitter le mode d'amélioration des séries refusées. Les deux autres choix vous font passer à la série suivante (s'il en reste à traiter) ; l'acceptation met à jour la table de projet et le fichier qui contient vos options de modélisation. En revanche si vous passez la série, son statut (refusé) n'est pas modifié dans le projet.

Dans l'assistant d'amélioration, un nouveau menu apparaît dans la barre des menus, ainsi qu'une nouvelle barre d'outils dont les éléments correspondent exactement aux éléments du menu.



La modélisation automatique (icône du robot) relance le traitement, de manière entièrement automatique. En fait, certaines des options du modèle sont modifiées, en répondant aux questions qui sont posées dans le dialogue correspondant à la modélisation personnalisée (icône de visage humain) en fonction des valeurs des statistiques obtenues dans le traitement initial. Cette amélioration automatique dispense donc le conjoncturiste de se poser les questions auxquelles le logiciel peut répondre.

Si vous choisissez le traitement personnalisé, il vous faudra répondre vous-même aux questions du dialogue qui sont autant d'indications (*hints*).



Demetra Customised Seasonal Adjustment Processing

SOME HINTS FOR QUICKLY IMPROVING THE MODELLING

The series behaviour changes sharply in this sample?
☐ ... perform the treatment on a shorted series sample

The series IS (logarithm) transformed but there is NO visible proportional relationship between the trend and the seasonal movements, or vice versa?
☐ ... modify the transformation specification

Only the number of outliers is too large?
☐ ... increase the significance level for outliers (see left below!)

Automatic ▾

Only the Ljung-Box/Box-Pierce statistics on squared residuals are significant?
☐ ... decrease the significance level for outliers (see left above!)

Only the Ljung-Box/Box-Pierce statistics on residuals are significant?
☐ ... check the trading day specification

Several of the problems mentioned above occur?
☐ ... customise more parameters at once

Look as well for practical events which could have influenced the series data, and construct regression variables able to explain these effects.

< Précédent Suivant > Annuler Aide

Demetra Customised Seasonal Adjustment Processing

SOME HINTS FOR QUICKLY IMPROVING THE MODELLING

The series behaviour changes sharply in this sample?
☐ ... stop the traitement here and split your sample manually

The series is (logarithm) transformed but there is no visible proportional relationship between the trend and the seasonal movements, or the other way around?
☐ ... modify the transformation specification

Only the number of outliers is too large?
☐ ... increase the significance level for outliers (see left below!)

Automatic ▾

Only the Ljung-Box/Box-Pierce statistics on squared residuals are significant?
☐ ... decrease the significance level for outliers (see left above!)

Only the Ljung-Box/Box-Pierce statistics on residuals are significant?
☐ ... check the trading day specification

Several of the problems mentioned above occur?
☐ ... customise more parameters at once

Look as well for practical events which could have influenced the series data, and construct regression variables able to explain these effects.

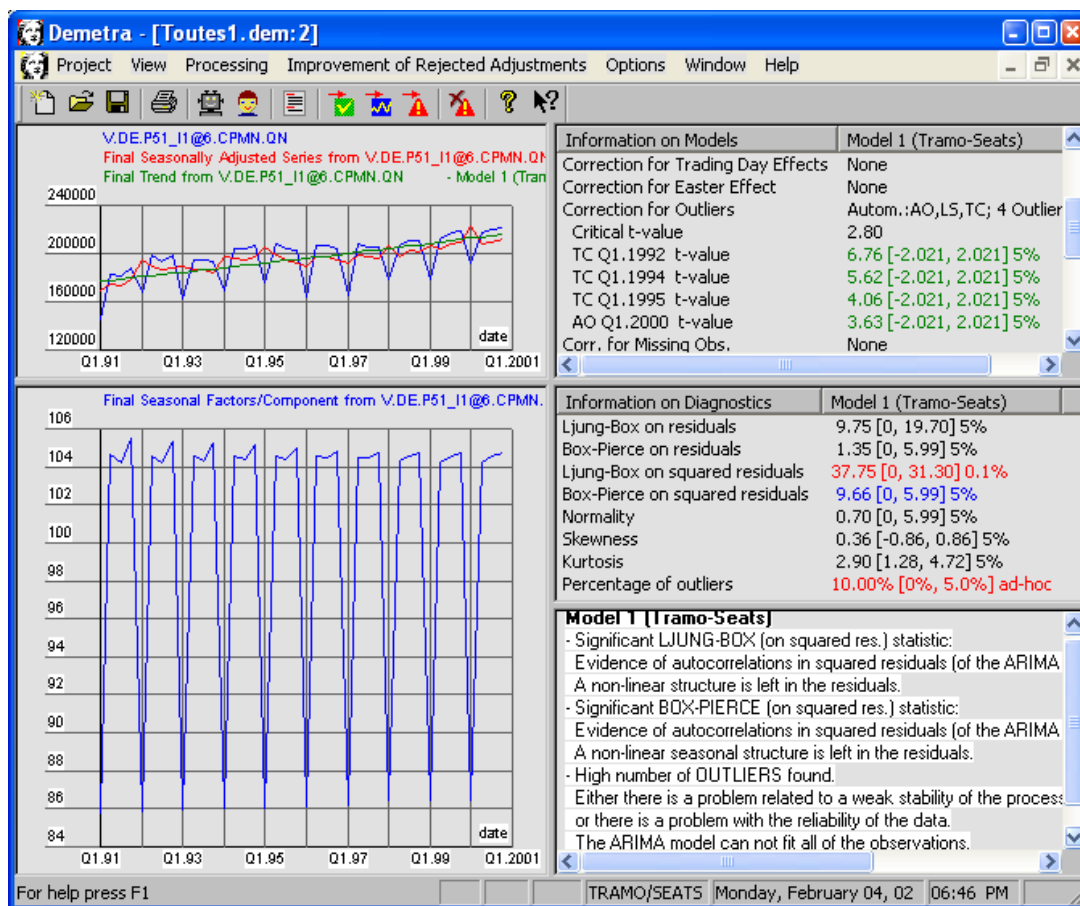
< Précédent Suivant > Annuler Aide

Q.1 La série présente-t-elle une brusque rupture ? La réponse à cette question se fait au vu des représentations graphiques. Nous avons abordé cette question plus haut. L'assistant va alors vous proposer de modifier la période de la modélisation.



Q.2 La série a-t-elle été transformée en logarithmes (ce qui suppose un modèle multiplicatif, mais les graphiques ne sont pas conformes à ce que l'on attend d'un tel modèle (voir ci-dessus, la présentation des schémas de décomposition) ? Il faut alors ne pas autoriser la transformation des données.

Q.3 La série présente-t-elle un nombre d'événements (*outliers*) trop important ? Ceci est l'indice que les algorithmes ont du mal à déterminer la tendance ; il est dans ce cas recommandé de modifier le seuil de détection des événements. Vous pouvez utiliser pour cela le menu déroulant situé à gauche. Celui-ci affiche par défaut un choix « automatique » et propose des valeurs allant de 2,8 à 4,2 par pas de 0,1. Pour diminuer le nombre d'*outliers*, il faut augmenter ce seuil. Il est possible de s'aider en analysant les statistiques fournies dans la fenêtre d'amélioration.



En haut à droite, figurent les événements détectés (ici, quatre événements, soit 10% des observations, valeur en rouge en bas à droite) ainsi qu'une mesure de l'importance de l'effet de chacun. Les événements sont classés par ordre décroissant d'importance (de 6,76 à 3,63). Ils ont été détectés en utilisant le choix automatique du seuil (ligne *Correction for outliers*) qui dans ce cas précis a fourni la valeur 2,80. Tous ces nombres sont des nombres sans dimension (sans unités), exprimés en unités d'écart type de la composante résiduelle.

Q.4 Les statistiques sur les carrés des résidus (Ljung-Box et Box-Pierce) sont-elles significatives ? Ceci indique qu'il reste une structure de corrélation dans les résidus qui ne ressemblent donc pas à un bruit blanc. Une des (nombreuses) façons d'obtenir une autocorrélation est la présence de point(s) extrême(s) qui n'auraient pas été détectés comme tels. Ceci explique la consigne de diminution du seuil de détection des valeurs extrêmes.

Q.5 Les mêmes statistiques sur les résidus sont-elles significatives ? Là encore, il reste donc une structure dans les résidus. Elle peut résulter, entre autres, de la non correction de l'effet de jours ouvrés ou, au contraire, d'une correction effectuée à tort.



Si une de ces questions, *et une seule*, aboutit à une réponse positive, le conjoncturiste doit agir sur les options indiquées avant de relancer le traitement. Pour les questions 2 et 5, le dialogue correspondant à l'option retenue succède donc à notre dialogue de suggestions. Pour les questions 3 et 4, il faut modifier la valeur du menu déroulant situé immédiatement à gauche. Le choix 1 interrompt le traitement de la série en cours et passe à la suivante.

Q.6 Si, au contraire, la réponse à plus d'une des questions est positive, les suggestions proposées ne sont plus applicables, l'utilisateur est donc laissé de nouveau à lui-même. Ce choix provoque donc l'affichage du dialogue de personnalisation du traitement que nous verrons un peu plus loin.

À l'issue de ces différentes tentatives d'amélioration de la série, l'utilisateur se trouve donc devant plusieurs modèles en concurrence. Il doit alors prendre une décision concernant le statut de la série dans le projet en cours :

- 4 accepter l'un des traitements,
- 4 transférer la série au module d'analyse détaillée pour poursuivre l'analyse en ayant accès à la totalité des paramètres des modèles,
- 4 ignorer la série, et donc la laisser dans le statut « refusé ».

Quelle que soit sa décision, il passe alors à la série suivante, jusqu'à épuisement des séries sélectionnées et passées à l'assistant d'amélioration.

Notons, pour les experts, que l'utilisation de l'assistant d'amélioration de l'ajustement permet à l'utilisateur expert d'obtenir les sorties complètes des méthodes utilisées et donc de disposer de l'ensemble des aides à l'interprétation et à la décision qui sont éditées par chacune des deux méthodes.

Log File for Processing 'Model 3 (Tramo-Seats)' of 'V.DE.P51_11@6.CPMN.QN' Page 1 / 25

TIME SERIES REGRESSION MODELS WITH ARIMA ERRORS, MISSING VALUES AND OUTLIERS.
BETA VERSION (*)

BY
VICTOR GOMEZ & AGUSTIN MARAVALL
with the programming assistance of G. CAPORELLI

(*) Copyright : V. GOMEZ, A. MARAVALL (1994,1996)

TRAMO compilation: 1999-01-28 23:24:15

SERIES TITLE=V.DE.P51_11@6.CPMN.QN

SINCE LONGER FORECAST FUNCTION IS REQUIRED
BY SEATS, NPRED CHANGED TO (8)

ORIGINAL SERIES

NUMBER OF OBSERVATIONS: 40

X 10.003

YEAR	1	2	3	4
1991	145.700	182.660	181.290	188.330
1992	167.520	198.370	193.990	198.620
1993	163.020	195.140	194.670	192.990
1994	169.080	204.610	203.450	207.160
1995	176.260	208.290	203.810	202.210
1996	163.530	205.690	206.570	203.710
1997	165.400	207.770	205.930	205.380
1998	178.690	206.490	211.170	210.540
1999	179.280	212.140	217.060	218.560
2000	192.380	217.430	219.490	221.370

DATES OF EASTER DURING THE REQUESTED TIME SPAN

YEAR	MONTH	DAY
1991	MARCH	31
1992	APRIL	19
1993	APRIL	11
1994	APRIL	3
1995	APRIL	16
1996	APRIL	7
1997	MARCH	30
1998	APRIL	12
1999	APRIL	4

sortie (log) Tramo-Seats (page 1 sur 25)



```

U. S. Department of Commerce, U. S. Census Bureau

X-12-ARIMA monthly seasonal adjustment Method,
Release Version 0.2.8

This method modifies the X-11 variant of Census Method II
by J. Shiskin A.H. Young and J.C. Musgrave of February, 1967,
and the X-11-ARIMA program based on the methodological research
developed by Estela Bee Dagum, Chief of the Seasonal Adjustment
and Time Series Staff of Statistics Canada, September, 1979.

Primary Programmers: Brian Monsell, Mark Otto

Series Title- CE,M,C91000U11B,UAL
Series Name- CE,M,C91000U11B,
Mon Feb  4 22:28:02 2002

-Period covered- 1st month,1972 to 5th month,2001
-Type of run - additive seasonal adjustment

-Sigma limits for graduating extreme values are 1.5 and 2.5 .
-3x3 moving average used in section 1 of each iteration,
3x5 moving average in section 2 of iterations B and C,
moving average for final seasonal factors chosen by Global MSR.
-Holiday adjustment factors applied directly to the final seasonally adjusted series
-Spectral plots generated for selected series
-Spectral plots generated for series starting in 1993.Jun
CE,M,C91000U11B,UAL
PAGE 1/28, SERIES CE,M,C

```

sortie (log) X12-Arima (page 1 sur 28)

Avec l'outil d'amélioration de l'ajustement, nous disposons donc d'un moyen de poursuivre, plus ou moins automatiquement, le traitement des séries à désaisonnaliser. Il fournit de nombreuses aides pour assister l'utilisateur dans ses décisions. Cependant, il présente l'inconvénient d'obliger le conjoncturiste à décider individuellement pour *chacune* des séries. Nous allons maintenant proposer une méthode de travail qui permette de continuer à bénéficier du traitement de masse.

II.3.2. Amélioration des modèles en traitement par lot

Notre méthode repose sur la possibilité de filtrer les séries pour n'afficher que les séries refusées, mais surtout sur la possibilité de trier la table des séries du projet en fonction de n'importe laquelle des colonnes de ladite table. Un premier exemple, démonstratif, utilise la dernière colonne de cette table : la colonne *Seasonality*. En cliquant sur l'en-tête de cette colonne, nous trions la table ; elle peut contenir trois valeurs distinctes :

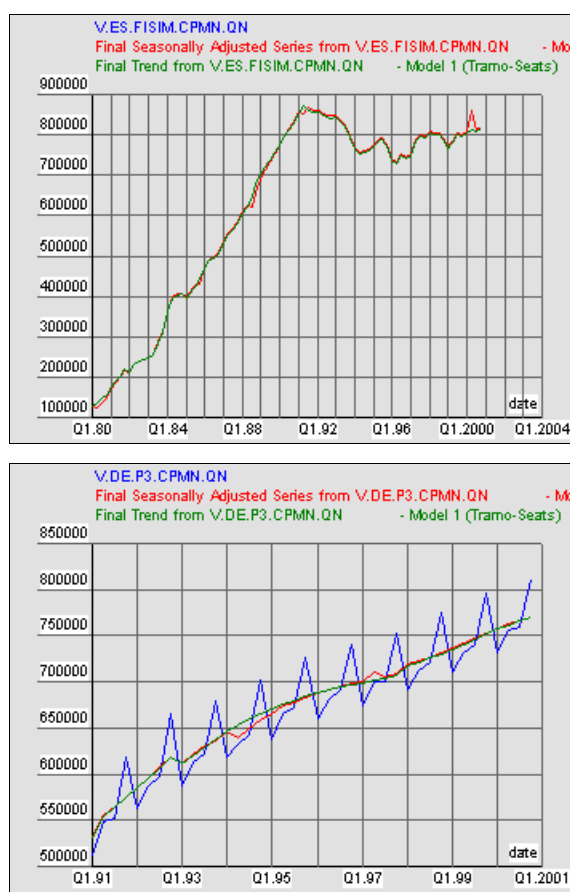
- 4 pour Tramo-Seats : *non seasonal model used, seasonal model used, to be checked*
- 4 ou pour X12-Arima : *not significant, probably present, significant*.

Explorant ensuite graphiquement les séries, il est alors facile de repérer celles qui, effectivement, ne nécessitent pas de désaisonnalisation.

Ø Trop de valeurs extrêmes

Retournons à nos séries à améliorer et reprenons certaines des indications fournies par l'assistant d'amélioration de l'ajustement. Commençons par la présence trop nombreuse d'événements ou valeurs extrêmes. Dans l'esprit de ce qui précède, il nous est facile de trier les séries en fonction de la colonne *Percentage of outliers*. Les séries à problème se retrouvent alors à la fin de la table.

Rappelons, encore une fois, qu'une démarche d'exploration graphique des séries ainsi regroupées peut toujours s'avérer profitable. Voici, par exemple deux séries présentant respectivement 13% et 15% d'*outliers* et donc rejetées par Demetra.



Dans le premier cas, la série n'est pas saisonnière. Dans le deuxième, la série est extrêmement régulière, quasi périodique, le modèle la décrit très bien ; la composante résiduelle est très faible et donc son écart-type très faible également. Tout point s'éloignant un peu du modèle se trouve de fait à une distance importante lorsqu'on l'évalue en unité d'écart-type.

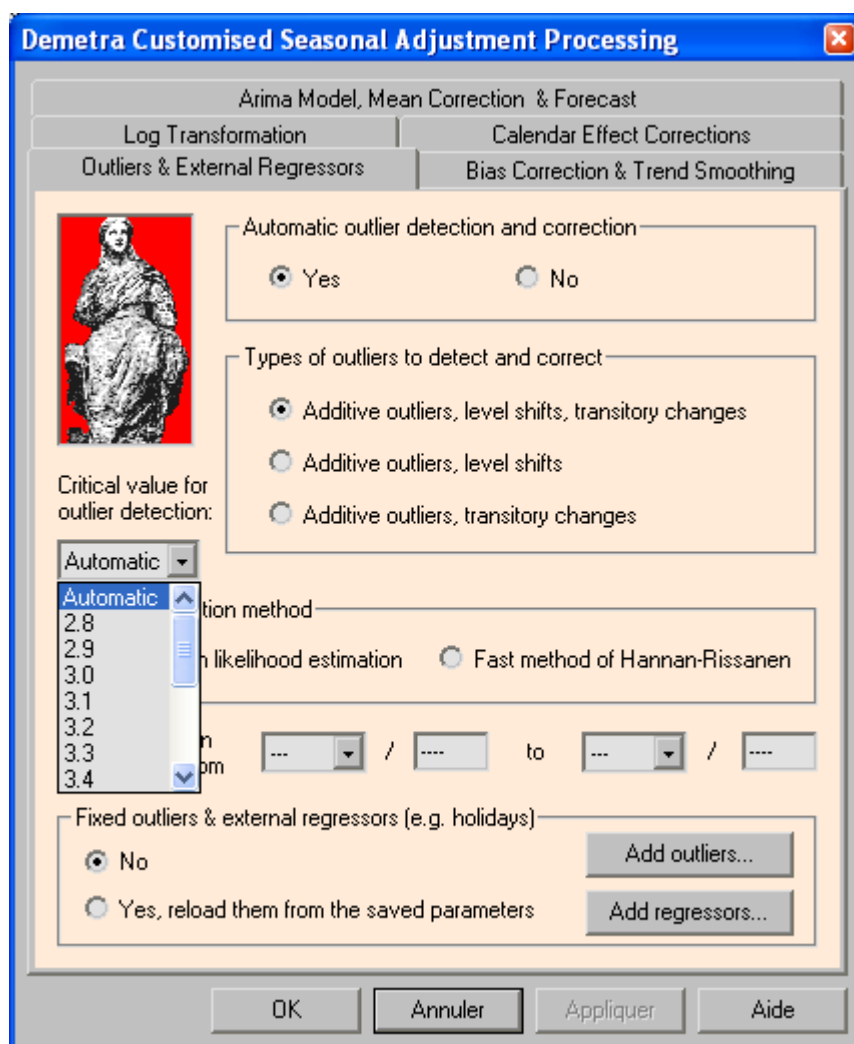
Ayant trié nos séries en fonction du nombre d'*outliers*, nous pouvons les sélectionner et effectuer, *globalement*, un traitement personnalisé adapté à ces séries, c'est-à-dire modifiant la détection et le traitement des *outliers* dans le modèle.

Dans le dialogue de traitement personnalisé, nous sélectionnons l'onglet correspondant où nous pourrions manipuler les options concernant les événements. Abordons les dans un ordre différent de celui du dialogue. Tout d'abord, nous y retrouvons le menu déroulant de fixation du seuil de détection des valeurs extrêmes ; c'est le même que celui que nous avons rencontré lors de l'assistant d'amélioration, il fonctionne donc de la même manière : augmenter le seuil pour diminuer le nombre d'événements détectés. On peut ainsi augmenter progressivement le seuil en le portant à 3,2, puis 3,7 et 4,2.

Ensuite, il est possible de restreindre le type d'événements à détecter automatiquement. En particulier, l'option sans les changements transitoires (TC) peut être intéressante à tester lorsque ceux-ci sont nombreux.

On peut également modifier la méthode de détection. La méthode rapide, comme son nom l'indique, n'est à utiliser que dans les cas où les performances de calcul doivent absolument être améliorées.

Il est possible de restreindre la période pendant laquelle Demetra recherche des événements. Attention ; dans un traitement par lot, cette restriction s'appliquera à toutes les séries du lot.



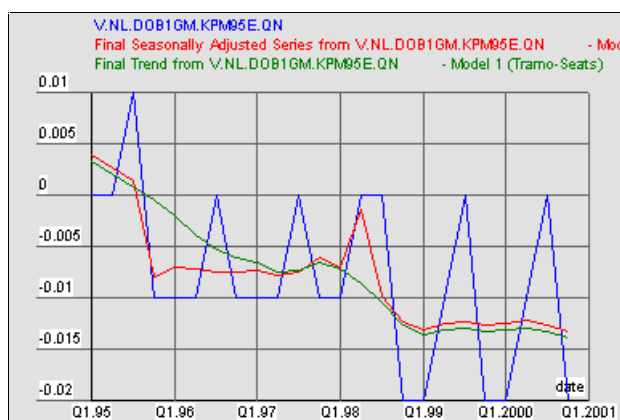
Enfin, dans les cas graves, et c'est la première option du dialogue, il est possible de désactiver complètement la détection et le traitement des valeurs extrêmes. Ce choix est à utiliser avec précaution. En effet, s'il permet d'évacuer complètement les *outliers* et donc de respecter les critères de qualité, il implique que les éventuels points extrêmes influent dans la détermination de la tendance. Dans ce cas, il est alors utile de spécifier soi-même les points extrêmes, ou d'introduire un régresseur externe (dernière zone du dialogue).

En modifiant progressivement les options dans l'ordre indiqué ci-dessus et en réexécutant l'ajustement à chaque fois sur le sous-ensemble des séries rejetées pour cause d'un nombre excessif d'événements, on a de fortes chances d'augmenter au fur et à mesure le nombre de modèles acceptés.

Ø Statistique de Ljung-Box sur les carrés des résidus

Continuons la logique de l'assistant d'amélioration, et considérons les séries pour lesquelles la statistique de Ljung-Box sur les carrés des résidus est trop élevée (significative). Attention, cette statistique n'est calculée que par la méthode Tramo-Seats ; X12-Arima ne le produit pas.

Nous trions donc notre table des séries en fonction de ce critère ; nous explorons graphiquement les séries dont la statistique apparaît en rouge (motif de rejet de l'ajustement). Cette opération permet souvent de mettre en évidence des séries particulières.



Sélectionnons ensuite les séries dont *seule* la statistique de Ljung-Box pour les carrés des résidus est significative, conformément aux indications de l'assistant d'amélioration et relançons l'ajustement personnalisé en diminuant le seuil de détection des *outliers*, par exemple en le fixant à la valeur minimale de 2,8.


Ø Statistique de Ljung-Box sur les résidus

Pour finir dans la logique de l'assistant d'amélioration, il nous reste à traiter du cas où seule la statistique de Ljung-Box sur les résidus est significative. Rappelons que ceci signifie que de l'information subsiste dans les résidus. Parmi les très nombreuses possibilités, l'assistant suggère la modification des paramètres de correction de jours ouvrés.

Pour tenir compte de cette suggestion dans un traitement par lot, il faut :

- 4 filtrer l'affichage pour ne conserver que les séries rejetées,
- 4 trier le tableau selon la colonne Statistique de Ljung-Box,
- 4 sélectionner parmi les séries du bas du tableau, dont la statistique s'affiche en rouge, celles qui partagent la même option de correction des jours ouvrés (cjo). Il peut être nécessaire d'effectuer une sélection discontinue (Ctrl-clic). On peut alors lancer un nouveau traitement par lot (option *default parameter for a new processing*) en imposant dans l'écran suivant un autre choix de cjo.





NEW AUTOMATIC SEASONAL ADJUSTMENT

Seasonal Adjustment Method

- ☒ Tramo/Seats
- ☐ X-12-Arima

Modelling Time Interval

/
to
/


☒ Otherwise reload from the saved parameters, if available

Type of Trading Day Effect to Test

- ☐ No trading day adjustment 0 regressors
- ☐ Working days (Monday to Friday): 1 regressor
- ☐ Working day (Monday to Friday) & leap-year: 2 regressors
- ☐ Trading day (Monday, Tuesday, ..., Saturday): 6 regressors
- ☒ Trading day (Mon, Tue, ..., Sat) & leap-year: 7 regressors

Country-Specific Holidays
☒ Allow reducing the number of trading day regressors

< Précédent
Suivant >
Annuler
Aide



NEW AUTOMATIC SEASONAL ADJUSTMENT

Seasonal Adjustment Method

- ☒ Tramo/Seats
- ☐ X-12-Arima

Type of Trading Day Effect to Test

- ☐ No trading day adjustment 0 regressors
- ☐ Working days (Monday to Friday): 1 regressor
- ☐ Working day (Monday to Friday) & leap-year: 2 regressors
- ☐ Trading day (Monday, Tuesday, ..., Saturday): 6 regressors
- ☒ Trading day (Mon, Tue, ..., Sat) & leap-year: 7 regressors

Country-Specific Holidays

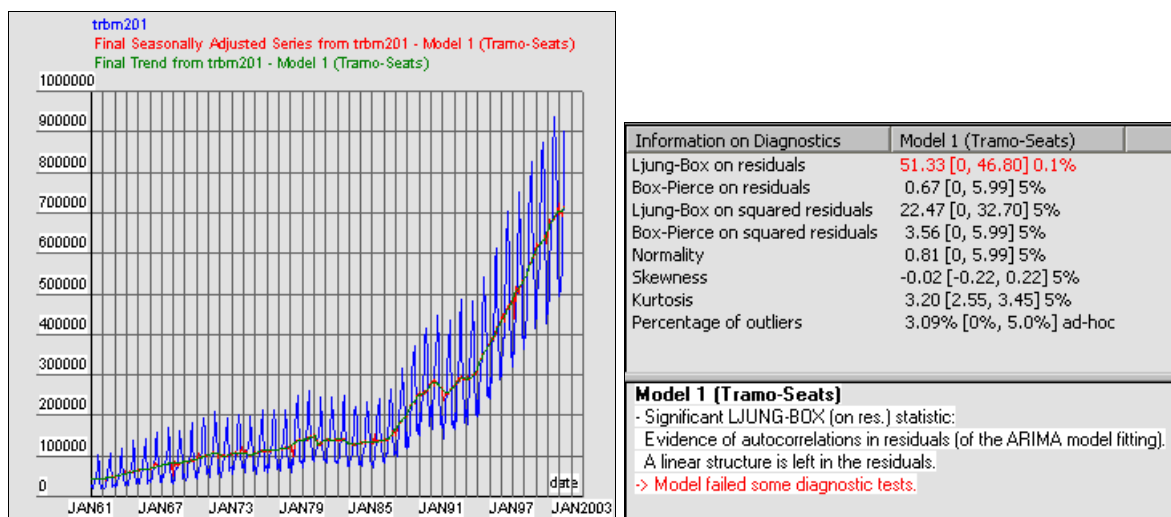
< Back
Suivant >
Cancel
Aide

Dans un but de rapidité, on peut d'ailleurs se cantonner aux deux choix extrêmes : pas de cjo contre correction avec 7 régresseurs.

Néanmoins, dans les exemples traités, il apparaît que ces essais ne permettent pas d'augmenter sensiblement le nombre de séries dont la modélisation est acceptée. Ici encore, une exploration



graphique montre que parmi les séries dont les statistiques sur les résidus sont significatives il en est souvent certaines qui présentent des ruptures : soit rupture de tendance, soit rupture de saisonnalité.



Ainsi, dans l'exemple ci-dessus, on voit clairement une rupture de tendance vers l'année 1986. On peut remarquer, d'ailleurs, que la série étudiée est particulièrement longue (40 ans) pour une désaisonnalisation. En la raccourcissant pour la faire démarrer en janvier 1986 (encore 20 ans), la série est acceptée par le traitement par défaut de Tramo/Seats.

Ø Autres parcours systématiques de la table du projet

Pour conclure cette partie qui suppose le traitement de (très) nombreuses séries, nous pouvons retenir qu'il peut être instructif de parcourir la table en visualisant graphiquement les séries en l'ordonnant en fonction des différentes colonnes de résultats statistiques. Nous en rappelons la liste dans le tableau ci-dessous, ainsi que les méthodes qui les produisent.

Nous avons parcouru les statistiques sur les résidus. On peut également se servir des statistiques portant sur la **distribution des résidus** (*normality*, *skewness*, *kurtosis*, soit normalité, asymétrie et aplatissement en français). Les valeurs de ces statistiques sont élevées lorsque la distribution des résidus s'éloigne de la distribution idéale postulée par la modélisation (loi normale). En les explorant, on pourra constater les éventuels accidents, ruptures ou changement de régime des séries pour lesquelles ces statistiques sont élevées.



Liste des statistiques de la table du projet

	Tramo-Seats	X12-Arima
Number	x	x
Name	x	x
Status of adjustment	x	x
Time span (n° of obs.)	x	x
Last time saved results	x	x
Arima model	x	x
Ljung-Box on residuals	x	x
Ljung-Box on squared residuals	x	
Box-Pierce on residuals	x	
Box-Pierce on squared residuals	x	
Normality	x	
Skewness	x	
Kurtosis	x	x
Forecast error		x
Percentage of outliers	x	x
Combined statistic Q		x
Transformation	x	x
Mean correction	x	x
Trading day effect	x	x
Easter effect	x	x
Outliers	x	x
Missing observations	x	x
Other regression effects	x	x
ARIMA decomposition	x	
X-11 decomposition		x
X-11 seasonal filter		x
X-11 trend filter		x
Seasonality	x	x

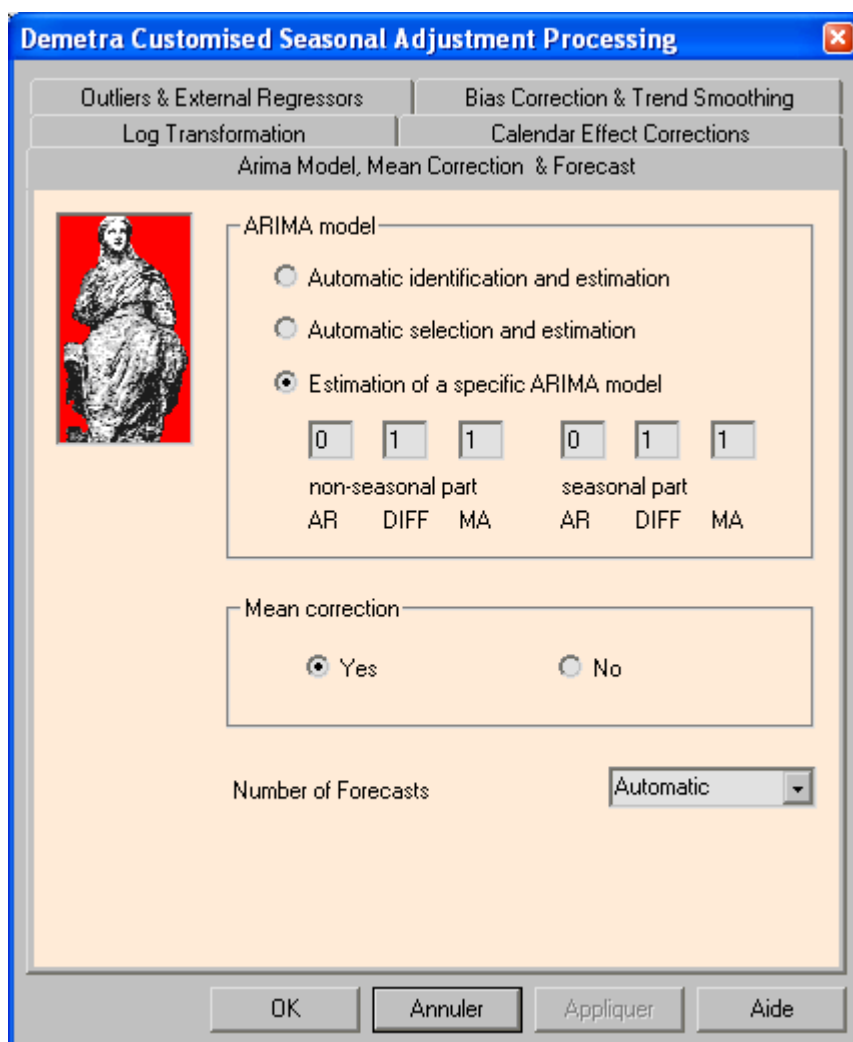
La colonne **ARIMA model** permet de trier les séries en fonction de l'ordre des modèles ARIMA sélectionnés. Lorsqu'on utilise la méthode Tramo-Seats, il est utile de s'intéresser aux modèles :

- 4 de type (0 0 0) pour la partie non saisonnière (premier triplet),
- 4 dont, à l'inverse l'ordre est élevé, ce qui indique que l'algorithme a eu du mal à déterminer la tendance.

Pour ces séries dont les modèles ont été rejetés, on peut essayer une dernière tentative de modélisation ARIMA en leur imposant le modèle le plus classique (dit *AIRLINE Model*). C'est celui que propose par défaut Demetra lorsque vous sélectionnez l'option : *Estimation of a specific ARIMA*



model. Il a pour ordre (0 1 1)(0 1 1) et est réputé bien décrire de nombreuses séries chronologiques courantes.



Demetra Customised Seasonal Adjustment Processing

Outliers & External Regressors | Bias Correction & Trend Smoothing
Log Transformation | Calendar Effect Corrections
Arima Model, Mean Correction & Forecast

ARIMA model

☐ Automatic identification and estimation
☐ Automatic selection and estimation
☒ Estimation of a specific ARIMA model

0 1 1 0 1 1
non-seasonal part seasonal part
AR DIFF MA AR DIFF MA

Mean correction

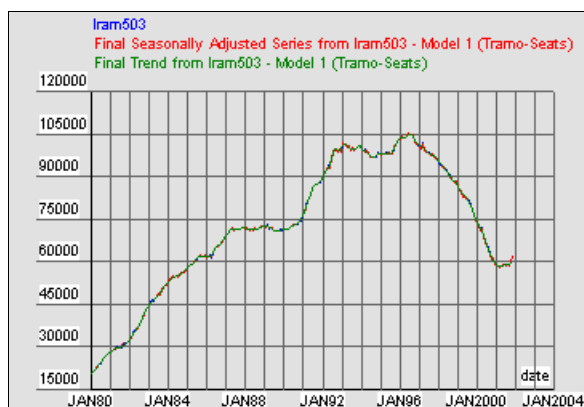
☒ Yes ☐ No

Number of Forecasts: Automatic

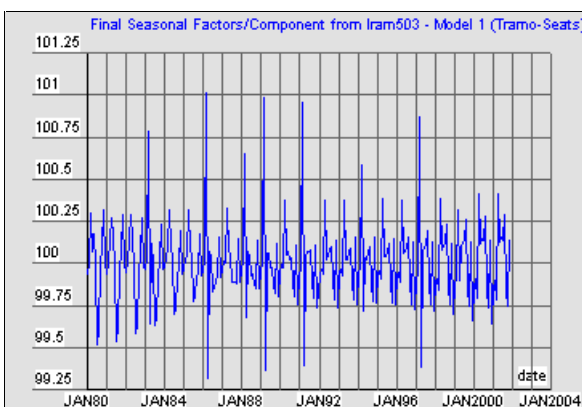
OK Annuler Appliquer Aide

Enfin, la dernière colonne de la table (**seasonality**) permet de regrouper les séries dont Demetra considère qu'elles ne présentent pas de composante saisonnière. On pourra s'assurer, visuellement qu'il en est bien ainsi. À l'inverse, certaines des séries que Demetra traite comme saisonnière peuvent, ne pas présenter de composante saisonnière très apparente.

série brute, tendance et cvs



composante saisonnière



Dans l'exemple ci-dessus, la composante saisonnière (multiplicative) est de l'ordre de $\pm 1\%$.

II.3.3. Système expert

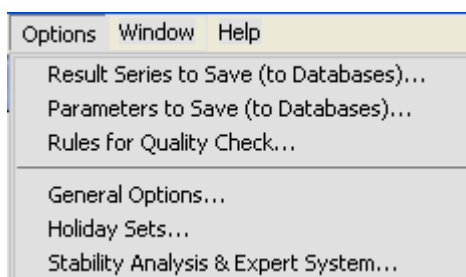
La version de mars 2002 de Demetra (version 2.0 Service Pack 1) a introduit un système expert pour automatiser l'amélioration des ajustements refusés. Lorsqu'une série est rejetée et que l'option est active le système expert va réexécuter toute une série d'ajustement en modifiant successivement les différents paramètres du modèle. Le manuel de l'utilisateur détaille les options qui sont modifiées.

Il est également possible d'autoriser Demetra à raccourcir la série par tranche de un an jusqu'à une période minimale d'analyse de quatre ans.

L'utilisation du système expert s'active en enfonçant l'icône correspondante dans la barre d'outil (un robot rouge à lunettes sur une balance)



ou en cochant la case dans le dialogue de paramétrage du système expert. C'est aussi dans ce dialogue que s'active l'option de raccourcissement automatique.





Parametrisation of Demetra's Stability Analysis & Expert System

Stability Analysis

☒ Systematically apply the Stability Analysis

Stepwise (by 1 period) cut model span at: ☐ Start ☒ End

Expert System

☒ Apply the Expert System on rejected adjustments

☐ Allow stepwise (by 1 year) cutting the model span at the start

☒ Give some information on the analysis progress and results

OK Cancel Restore Defaults

Total Model Quality

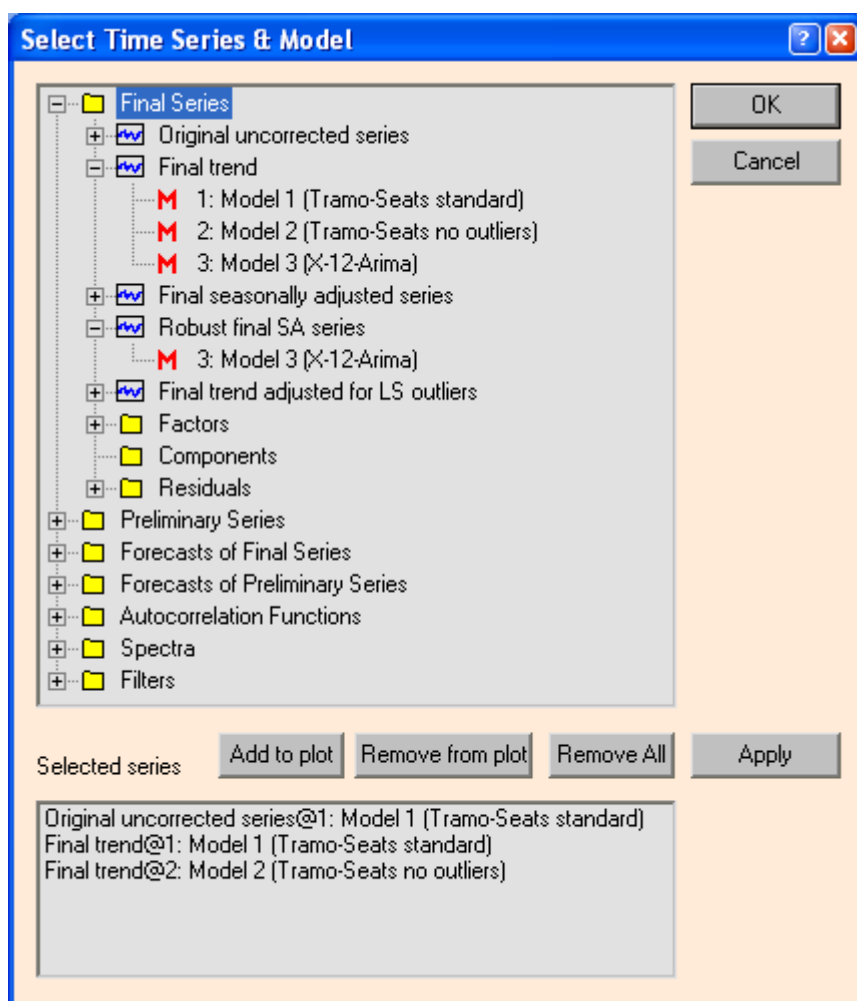
Importance of history	0
Exponential penalty	1
Weight of Ljung-Box	1
Weight of Box-Pierce	1
Weight of Ljung-Box (Sq.)	1
Weight of Box-Pierce (Sq.)	1
Weight of normality	1
Weight of skewness	1
Weight of kurtosis	1
Weight of forecast error	1
Weight of n° of outliers	1
Weight of X11-Q statistic	1

II.3.4. Traitement dans le module d'analyse détaillée

Si vous disposez du temps nécessaire, le passage dans le module d'analyse détaillée peut vous être utile. Il permet de stocker les paramètres et les résultats de différents essais de modèle sur une même série. En particulier, il autorise à tester les deux grandes familles de méthodes (Tramo-Seats et X12-ARIMA), ce que ne permet pas l'assistant d'amélioration des séries refusées. Il donne accès à l'ensemble des paramètres de ces deux méthodes : les réglages déjà accessibles dans le traitement automatisé sont de nouveau présents, mais dans une interface totalement différente et d'autres paramètres, encore plus techniques deviennent accessibles.

Enfin, et c'est peut-être l'un de ses principaux intérêts, le module d'analyse détaillée propose un outil de comparaison graphique extrêmement pratique. Le menu ne s'active que lorsque un modèle au moins a été exécuté dans le module d'analyse détaillée. L'outil de comparaison graphique affiche une fenêtre à quatre volets. Dans chacun de ces volets, il est possible de définir un graphique composé de une ou plusieurs séries. Les séries représentées dans un même volet doivent être compatibles entre elles (même échelle).

Le dialogue de définition d'un graphique se présente comme ci-dessous. Chaque icône de dossier contient une ou plusieurs séries si le nœud de l'arbre qui la porte est muni du signe plus. La série est représentée par une petite courbe bleue (*line chart*) ; en double cliquant sur cet icône, on déploiera alors les courbes correspondant aux différents modèles testés dans le module d'analyse détaillée.

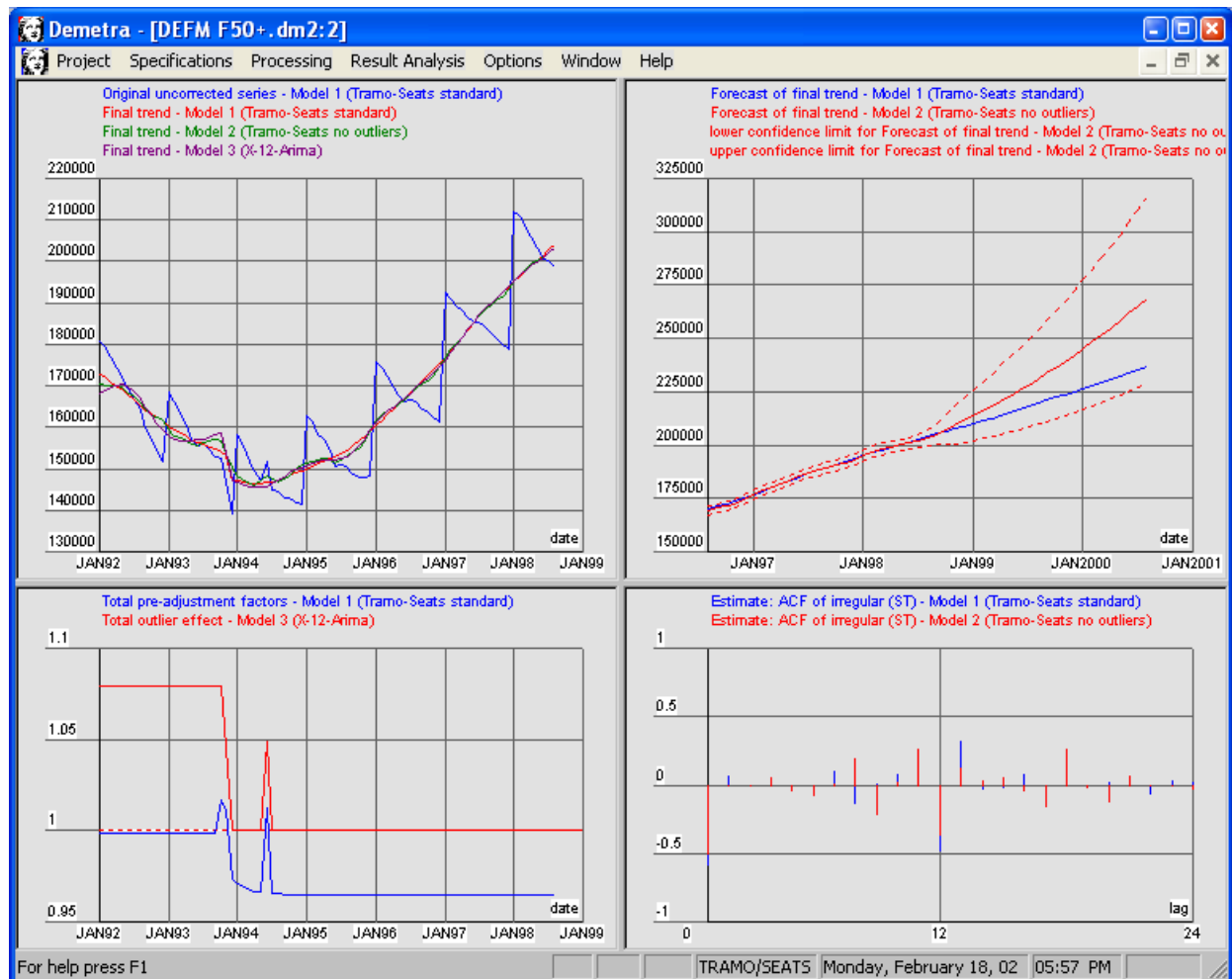


Attention, Tramo-Seats et X12-ARIMA ne produisent pas exactement les mêmes ensembles de séries ; il peut donc se produire que tous les modèles ne soient pas listés dans les choix possibles d'un des types de série à représenter.

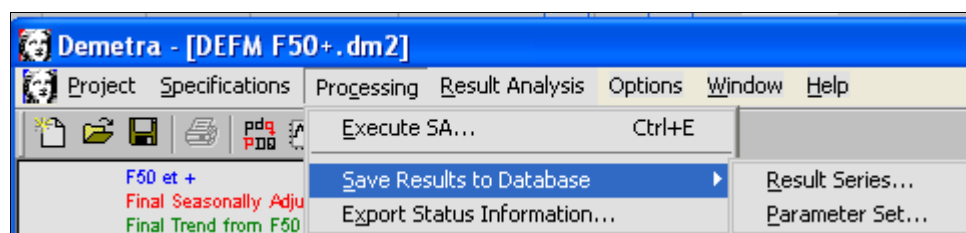
Dans l'exemple ci-dessous, on a représenté, en haut à gauche, la série originale ainsi que les tendances estimées par trois modèles différents, à droite les prévisions de la tendance fournies par les deux modèles Tramo-Seats.

En bas à gauche, on peut comparer les événements détectés par les deux modèles où l'on a activé la détection automatique des *outliers*. On peut remarquer que Tramo-Seats et X12-ARIMA ne calent pas leur correction de la même manière, mais qu'ils détectent les mêmes événements à l'exception de ceux de type TC dont la recherche n'est pas prévue dans l'algorithme de X12-ARIMA.

Enfin, en bas à droite, on trouve la représentation de la fonction d'autocorrélation des résidus des modèles Tramo-Seats. Ce graphique est l'un des outils indispensables à ceux qui souhaitent affiner la modélisation ARIMA.



Notons, pour finir, qu'il est possible de transférer les résultats d'un des modèles analysés dans le module détaillé au projet contenant la série ainsi analysée. Il en va de même du jeu de paramètres retenus dans ce modèle.



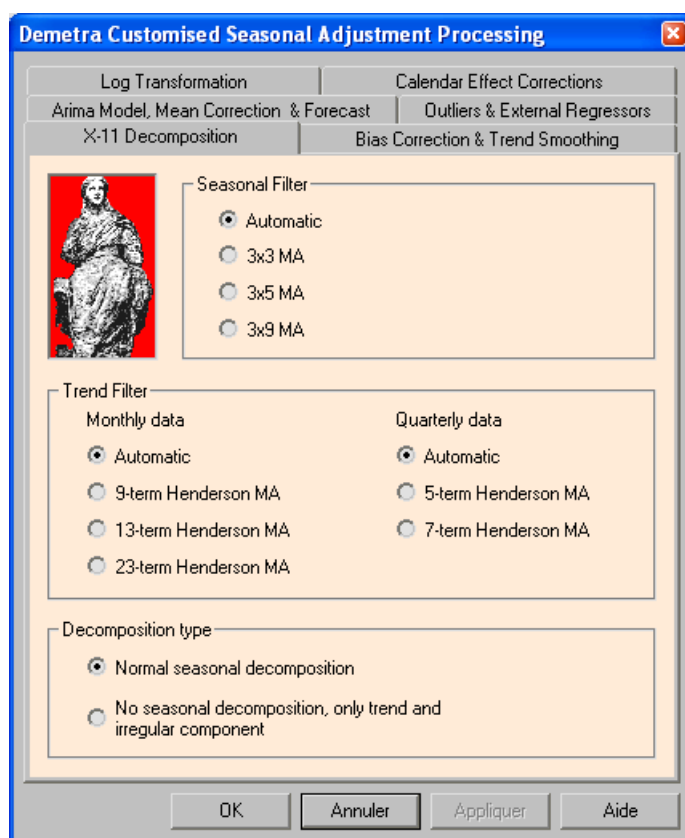
Ceci permet de stocker les informations du modèle sélectionné à l'aide du module d'analyse détaillée global dans la base de données d'où est issue la série étudiée. Ainsi, dans Excel, les onglets du classeur contenant la source seront mis à jour ou créés pour conserver les composantes produites par le modèle, telles que définies dans les options de stockage.

II.4. Spécificités de l'utilisation de X12-Arima

II.4.1. Paramétrage de X12-Arima

Si vous avez choisi la méthode X12-Arima, vous obtiendrez un dialogue de personnalisation de l'ajustement qui est en grande partie similaire à celui de la méthode Tramo-Seats. Quelques différences sont néanmoins à signaler : une partie des choix dans Tramo-Seats ne sont pas présents. Notons, par exemple, l'impossibilité d'introduire une variable d'intervention constituée d'une séquence de uns (bouton *Add outliers...* dans l'onglet *Outliers & External regressors*) ou encore un choix plus restreint pour la correction de biais de la série cvs.

En revanche, un onglet supplémentaire (*X11 Decomposition*) apparaît dans le dialogue de personnalisation. Il s'agit essentiellement du choix des ordres des différentes moyennes mobiles sur lesquelles repose la méthode X12.



Ø Trend filter

Vous pouvez choisir l'ordre, c'est-à-dire le nombre de termes, de la moyenne mobile utilisée pour estimer la tendance. Il s'agit de moyennes mobiles dites de Henderson qui présentent de bonnes propriétés de lissage. Plus le nombre de termes est élevé (9, 13, 23 mois ou 5 ou 7 trimestres), plus la tendance sera lisse. Par défaut, X12 choisit automatiquement l'ordre en fonction de la part relative de l'irrégulier dans les variations de la série.



Ø Seasonal filter

Il s'agit de l'ordre des moyennes mobiles estimant les coefficients saisonniers à partir des écarts à la tendance. Ce paramètre contrôle les possibilités d'évolution, dans le temps, des coefficients saisonniers. Sa plus petite valeur, 3x3, soit une moyenne des écarts à la tendance utilisant le même mois de 5 années consécutives pour estimer le coefficient saisonnier du mois de l'année médiane, correspond à une souplesse accrue sur les coefficients. À l'opposé, 3x5 (7 années) puis 3x9 (11 années) imposent une évolution plus lente des coefficients, c'est-à-dire limite les possibilités de déformation de la saisonnalité sur l'ensemble de la période.

Bien entendu, plus l'ordre de cette moyenne mobile est élevé et plus se pose avec acuité le problème des extrémités de série et donc la nécessité de disposer d'un « bon » modèle Arima pour compléter les extrémités.

Par défaut, X12 sélectionne l'option qu'il estime la plus adaptée.

Ø Decomposition type

Le choix est donné de désactiver la désaisonnalisation par X12 : elle doit être utilisée sur des séries non saisonnières ou déjà désaisonnalisées. Elle permet dans ce cas d'obtenir une estimation de tendance ainsi que diverses statistiques calculées par X12.

II.4.2. Exploration des résultats dans le cas de X12-Arima

Si vous utilisez X12-Arima, la liste des statistiques disponibles dans le tableau du projet est différente de celles fournies par Tramo-Seats. Moins de statistiques sont calculées sur les résidus. De même, la liste des modèles Arima testés est plus limitée ; dans cette colonne peut figurer la mention *none of the models were chosen*, lorsqu'aucun des modèles possibles n'a été sélectionné. Cette situation entraîne le rejet du modèle ; c'est une des options par défaut du contrôle de qualité de l'ajustement.

En revanche, dans les colonnes renseignées par X12, apparaît une colonne erreur de prévision (*forecast error*). Il s'agit, en fait, de « rétroprévisions » : X12 utilise le modèle Arima retenu pour fabriquer des « prévisions » des valeurs observées pendant la dernière année à partir des données des années précédentes. L'écart est mesuré en pourcentage ; s'il est trop important, la méthode n'utilisera pas l'extrapolation de la série par le modèle Arima pour compléter les données. Cette information est signalée dans la colonne « *X11-decomposition* » par l'indication *without ARIMA forecasts*.

Les colonnes *X-11 seasonal filter* et *X-11 trend filter* indiquent l'ordre des moyennes mobiles retenues (voir paragraphe précédent).

Enfin, la statistique combinée Q (*combined statistic Q*) est un indicateur combinant diverses statistiques particulières liées essentiellement à la stabilité des différentes composantes du modèle au cours de la période étudiée. Sa valeur est comprise entre 0 et 3 (sans dimension) et ne devrait pas être supérieure à 1.

II.5. Stratégie de mise à jour et de révision

Dans les parties précédentes, nous avons décrit l'utilisation de Demetra pour la mise en place initiale des traitements de désaisonnalisation. La démarche présentée, appuyée essentiellement sur l'outil d'analyse automatique et ses facilités de traitement par lot a permis de statuer sur l'ensemble des séries contenues dans le projet. Si le conjoncturiste n'a pas modifié les cases à cocher du dialogue de lancement du traitement, la définition des modèles retenus a été stockée dans le fichier de projet ;



ceci comprend le choix de la méthode appliquée (Tramo-Seats ou X12-Arima), la spécification des traitements préalables (*outliers*, transformation des données,...), l'identification des modèles Arima utilisés (les ordres des différentes parties) et l'estimation des coefficients de ces modèles. Enfin, l'utilisation de ce modèle a produit les différentes séries résultant de la décomposition saisonnière (tendance, cvs, irrégulier, etc.)

Par la suite, de nouvelles données vont être produites périodiquement, par exemple mensuellement, le conjoncturiste doit alors fournir des données corrigées des variations saisonnières et, éventuellement, proposer un commentaire de l'évolution récente. Il doit donc désaisonnaliser de nouveau les séries dont il dispose, complétées des points nouvellement acquis. Doit-il parcourir toute la démarche initiale ?

La réponse à cette question n'est pas simple. Elle est directement liée à l'organisation de son travail, résultant, d'une part de ses choix personnels, mais aussi de ceux de son environnement et, en tout premier lieu, de ceux de l'institution pour laquelle il travaille. De plus, la simple application d'un même modèle à des données légèrement différentes (une observation de plus !) peut entraîner une modification des valeurs passées des composantes. Ainsi, en prenant un exemple utilisant X12-Arima, l'apparition du point supplémentaire modifie (légèrement) les prévisions des observations utilisées pour estimer la tendance ; et ceci, même en n'ayant rien changé à la spécification du modèle. La tendance étant légèrement différente, l'équation comptable du modèle de décomposition implique que la totalité des autres composantes peut elle aussi varier.

Ainsi l'arrivée d'une donnée à la date t permet-elle de calculer la valeur cvs à la date t , mais elle peut aussi provoquer la révision des données cvs concernant les dates $t - 1$, $t - 2$, ...

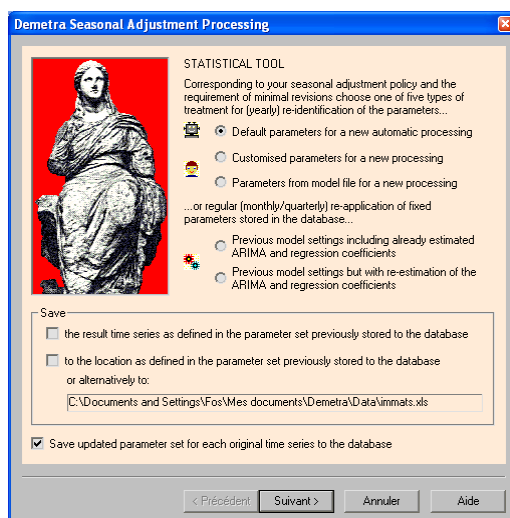
L'ampleur des révisions susceptibles de se produire dépend, bien évidemment, des modifications éventuelles apportées au modèle (réestimation des coefficients, réidentification du modèle, modification des traitements préalables, voire changement de méthode !) Elles sont toujours présentes, même en l'absence de toute modification.

Dans l'organisation des travaux de désaisonnalisation, il peut être utile de fixer une limite au délai au delà duquel, on ne corrige plus rétrospectivement les séries publiées.

Le schéma suivant résume une organisation possible des travaux de désaisonnalisation. Elle correspond à une application périodique du modèle, *sans réestimation* des coefficients, pour produire les composantes du modèle saisonnier. Cette organisation est fréquemment rencontrée mais ne doit pas être considérée comme normative.

		politique de l'institution	premier ajustement	nouvelles données	révision	refonte
périodicité			<i>une fois</i>	<i>mensuelle</i>	<i>annuelle ?</i>	
décisions	choix de la méthode (TS ou X12)	x ?				
	choix du modèle					
	options de prétraitement		x			x
	identification du modèle		x			x
	estimation des coefficients		x		x	x
	application du modèle		x	x	x	x
options d'ajustement		1, 2, (3 ?)		4	5	1, 2, (3 ?)

Les options d'ajustement indiquées dans le bas de ce tableau correspondent aux différents boutons radio du dialogue de lancement du processus de désaisonnalisation dans Demetra.



Une autre possibilité, parfois encore utilisée, consistait à n'estimer un modèle de désaisonnalisation qu'une seule fois par an afin de produire, non pas les données cvs, mais les prévisions des coefficients saisonniers pour les 12 mois à venir. Il suffisait, par la suite, d'appliquer ces coefficients aux données brutes pour obtenir, à l'aide d'une seule opération arithmétique, la valeur corrigée des variations saisonnières.

II.6. Quelques situations typiques

II.6.1. Changement de saisonnalité sur la période d'étude

Les méthodes utilisées par Demetra autorisent une modification lente de l'effet saisonnier sur l'ensemble de la période étudiée, c'est à dire une déformation progressive de la structure des coefficients saisonniers du modèle. Cependant, certaines séries peuvent présenter une rupture brusque de la saisonnalité, liée par exemple à une modification de la définition de la variable suivie, ou de son mode de calcul ou encore à un changement de dispositions réglementaires ou législatives. Demetra ne prévoit pas la détection de telles ruptures et les modèles retenus pour des séries en présentant peuvent être acceptés sans que le problème ait été détecté avec des conséquences importantes quant à la fiabilité de la décomposition proposée, en particulier autour du point de rupture.

C'est donc à l'utilisateur de chercher de telles situations au cours de son exploration visuelle.

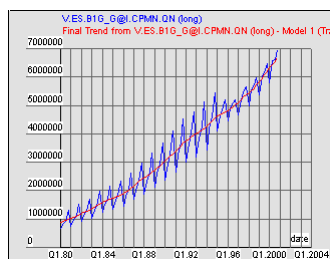
Ø **Modification de la période d'étude**

Voici une série affectée par un changement brusque de saisonnalité (au premier trimestre 1995). Nous allons l'étudier sur trois périodes différentes : l'ensemble de la série (21 ans, 84 observations), sur 11 années et enfin sur la dernière période après la rupture de saisonnalité de début 1995 (6 ans, 24 observations). Tous les résultats présentés le sont avec les options standards de Tramo-Seats pour un nouveau traitement.



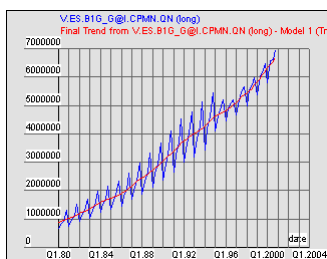
série brute et tendance

1980-2000



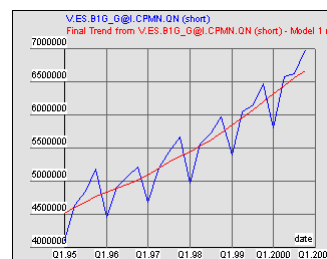
refusé

1990-2000



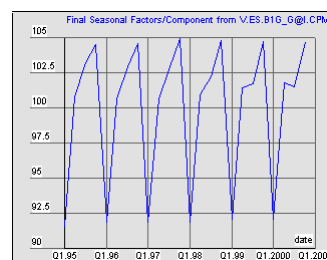
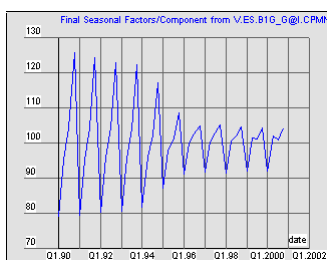
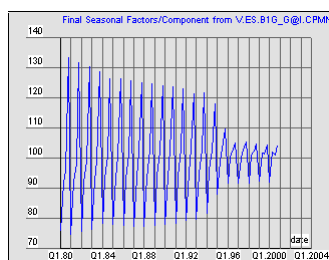
accepté

1995-2000

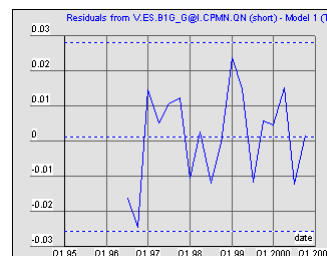
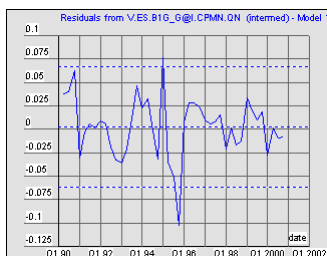
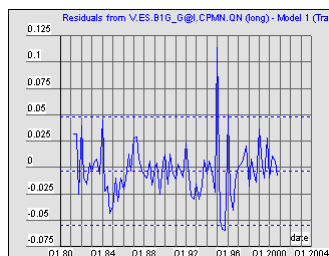


accepté

composante saisonnière



résidus



Dans la deuxième option, une statistique sur les résidus est significative à 5%, liée, sans doute à la configuration des résidus lors de l'année de transition, 1995. Mais cela ne suffit pas pour rejeter le modèle.

Quel est le meilleur choix ? Autrement dit quel intérêt a-t-on à conserver l'information du passé, même lointain.

Par ailleurs, les ordres de la partie non saisonnière des modèles ARIMA sont les suivants : (1 1 3) pour la plus longue période d'étude et (0 1 1) pour les deux autres. L'argument dit de « parcimonie » favorise, à résultats équivalents, les modèles dont le degré de complexité est le plus faible. Donc, là encore, même si l'on pourrait améliorer le modèle de la série longue pour finir par le rendre acceptable, le jeu n'en vaudrait sans doute pas la chandelle.

Donc, dans une approche pragmatique, il semblerait que le troisième choix, la période la plus courte, est celui qui permette de produire des estimations actuelles de la tendance. Le comportement de la série a clairement changé et, de plus, on dispose de suffisamment d'observations sous le nouveau régime de fonctionnement pour pouvoir estimer un modèle.

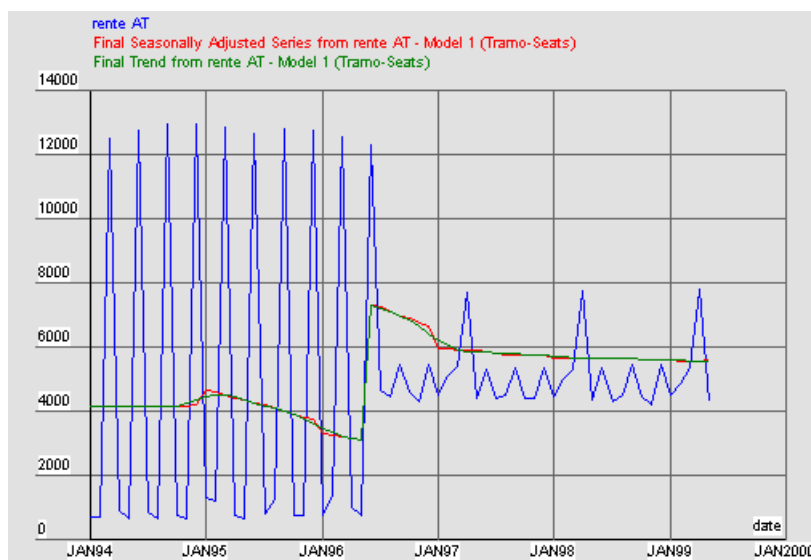
Dans cette approche, on réalisera de manière séparée les décompositions saisonnières pour les deux périodes :



- pour la période 1980-1995 (avant rupture)
- pour la période après la rupture de 1995.

Si la rupture de saisonnalité est trop récente, le conjoncturiste ne disposera pas d'un nombre d'observations suffisant pour pouvoir estimer correctement les paramètres de la nouvelle saisonnalité. Toute analyse conjoncturelle de la série sera alors délicate. D'ailleurs, il ne pourra pas faire tourner Demetra sur la période la plus récente si le nombre de points est insuffisant.

Ø Intégration du changement de saisonnalité dans le modèle



Au vu de la représentation graphique, il semblerait qu'une modification du comportement saisonnier de la série ait eu lieu en juillet 1996. Malgré cela, le modèle par défaut est accepté par Demetra : toutes les statistiques de qualité de l'ajustement sont bonnes ! Un tel constat ne peut que justifier le caractère **incontournable** de l'exploration graphique, même rapide, des données ! L'existence d'un problème grave sur la série est perçu instantanément par l'utilisateur ; l'étude un peu plus détaillée des résultats le persuadera rapidement de l'inadéquation du modèle retenu, ne serait-ce que par la transformation en logarithmes...

Dans des cas de même nature, suivant l'ampleur du changement intervenu et la longueur respective des séquences avant et après changement de saisonnalité, le modèle par défaut peut être accepté.

Dans l'exemple présenté ici, qui provient de données réelles, on connaît précisément la modification intervenue : il s'agit d'une prestation particulière versée initialement en fin de trimestre. À partir de mi-96, il a été décidé de verser un acompte mensuel et d'effectuer la régularisation des montants versés en avril de chaque année.

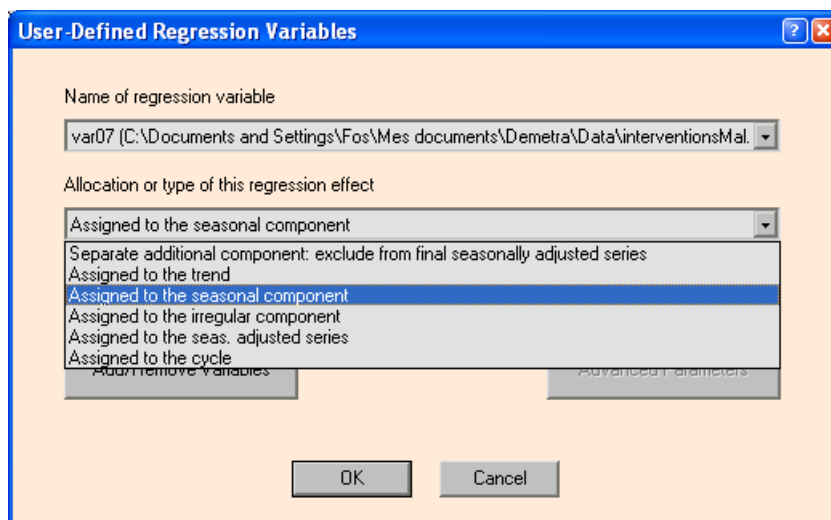
Ayant identifié la difficulté, quelles solutions pouvons nous apporter ?

La solution la plus simple et la plus radicale, consisterait à scinder la série en deux parties. Cela suppose, bien sûr, que la deuxième partie, après rupture, dispose de suffisamment d'observations pour pouvoir estimer un modèle.

Une autre solution, beaucoup plus lourde à mettre en œuvre consiste à introduire des variables externes dont l'effet viendrait porter sur la composante saisonnière. Il nous faut créer 11 variables d'intervention, autant que de saisons dans l'année moins un. Les valeurs prises par toutes ces variables sont 0 jusqu'à juin 1996 inclus. À partir de cette date, elles pourront prendre trois valeurs : +1, 0 et -1. Ainsi, nous allons créer une première variable, par exemple var01 correspondant au mois de janvier ; elle vaudra +1 pour tous les mois de janvier postérieurs à juin 1996, -1 pour les mois de



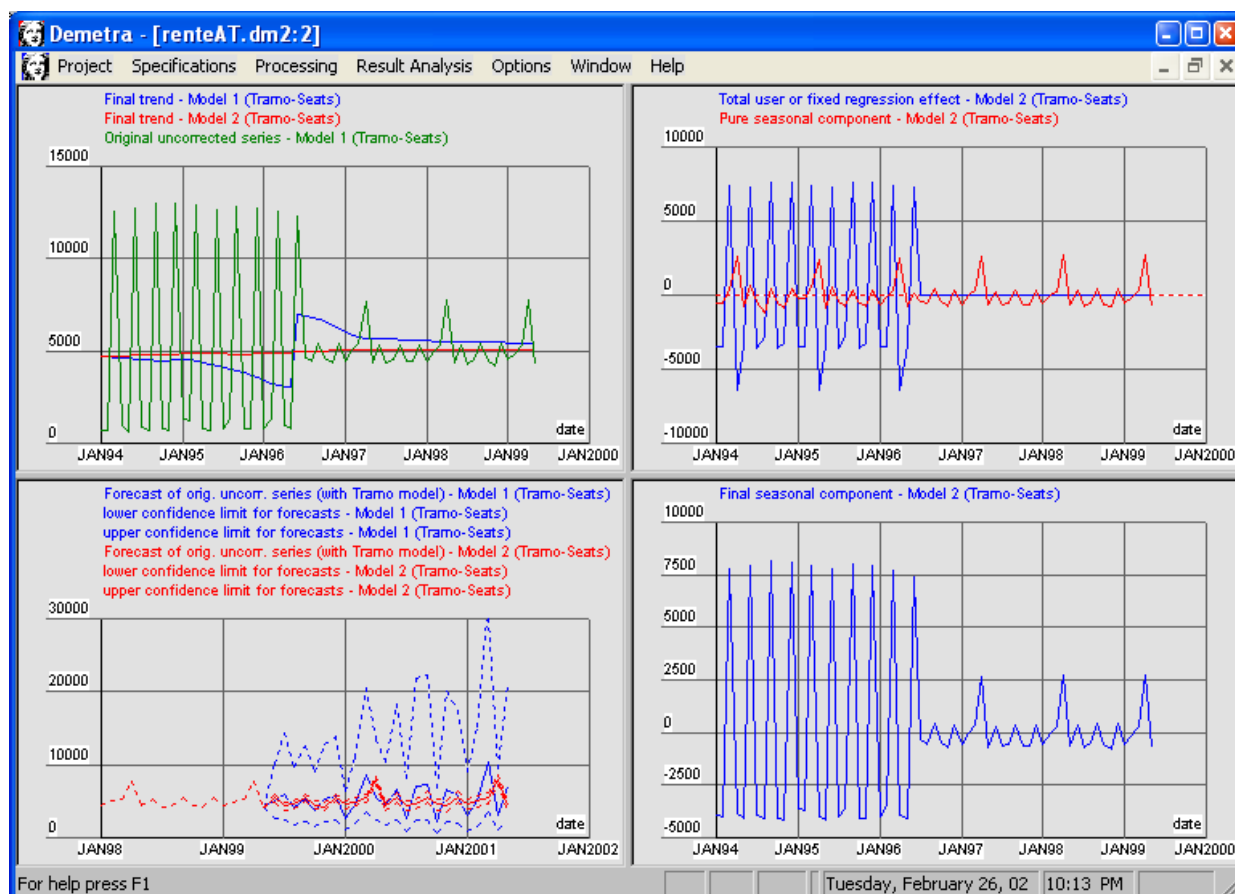
décembre de 1996 et années suivantes et 0 partout ailleurs. Nous créons de même des variables var02,..., var11 correspondant aux mois de février à novembre. Il n'y a pas de variable var12 pour le mois de décembre, en effet, la structure très particulière de nos onze variables garantit que la somme des variations de saisonnalité provoquée par la rupture soit nulle.



Nous introduirons ensuite ces variables en tant que régresseurs externes, utilisant ainsi la méthode introduite dans la partie précédente. Nous affecterons toutes ces variables à la composante saisonnière. Chacune d'entre elles fournit l'estimation de la variation du coefficient saisonnier résultant de la rupture de saisonnalité. Pour pouvoir réaliser une prévision de la série modélisée, il est nécessaire de prolonger la structure que nous venons de décrire sur l'ensemble de la période à prévoir.

Dans le paramétrage que nous venons de proposer, la situation de référence est constituée par la partie *précédant* la rupture. Dans l'optique de désaisonnalisation à des fins de conjoncture ou de prévision, il est peut être préférable de réaliser l'inverse : considérer la situation actuelle (après la rupture) comme la référence et reparamétrer nos variables externes pour qu'elles estiment l'écart du passé par rapport au présent. Dans cette approche, ce sont les valeurs passées des régresseurs qui sont constituées des valeurs +1, 0 ou -1 ; les valeurs récentes et, surtout, futures sont toutes égales à 0. Cette deuxième approche permet, à terme, lorsqu'on dispose d'un nombre de points suffisants, d'abandonner la partie la plus ancienne de la série sans avoir à reprendre entièrement le paramétrage.

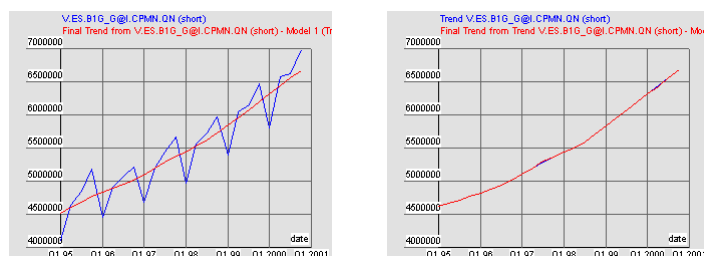
C'est ce choix que nous présentons dans la figure suivante, obtenue à l'aide de l'outil de comparaison graphique. Dans la partie gauche, figure, en haut, la représentation de la série brute et les deux tendances calculées l'une par le modèle par défaut, l'autre après introduction des régresseurs ; en bas, les prévisions fournies par ces deux modèles. Les données jusqu'en mai 1999 sont des données observées, les prévisions véritables débutent après ; elles sont accompagnées d'intervalles de confiance sur les prévisions fournies. On peut constater que les prévisions fournies par modèle avec régresseurs sont d'une part plus conformes à ce que l'on attend d'une série dont le comportement est aussi régulier, mais également beaucoup plus précises.



La partie droite décrit la composante saisonnière telle qu'elle est estimée par le modèle avec rupture de saisonnalité : en bas, la composante « finale », en haut, la façon de l'obtenir. La série en rouge représente la composante saisonnière « pure », c'est-à-dire sans tenir compte des régresseurs ; avec le paramétrage retenu, elle est donc estimée sur la partie récente de la série. La série en bleu résume l'impact des 11 régresseurs, sa valeur est nulle sur la fin de la période. La composante finale est la somme de ces deux séries.

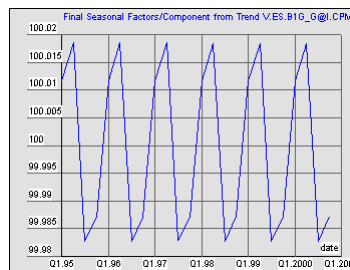
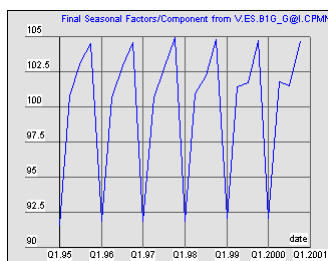
II.6.2. Non idempotence : saisonnalité dans la tendance

Le terme idempotence, peu courant, signifie que l'application d'une transformation sur des données sur lesquelles la transformation a déjà été appliquée une première fois donne un résultat identique à la première transformation. Dans notre cas, si on désaisonnalise la tendance obtenue après une désaisonnalisation, on devrait obtenir la même série. Il n'en est rien. Partant de l'exemple précédent en retenant la période la plus courte, on peut désaisonnaliser la tendance (graphique représenté ci-dessous à droite) ou la série cvs (non représenté), on trouve des résultats différents.





composante saisonnière



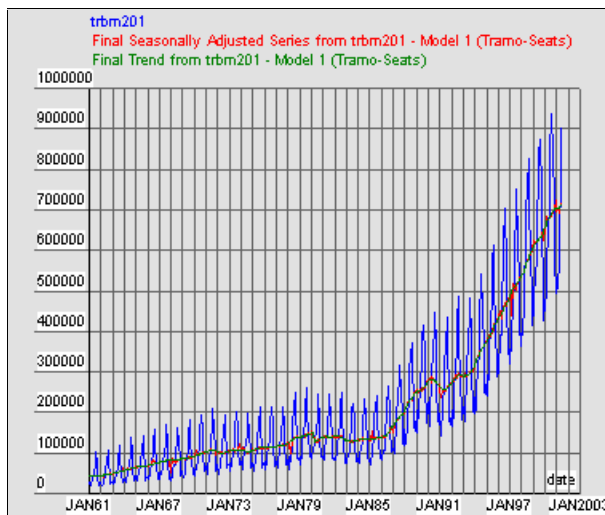
En particulier, on retrouve, très bizarrement, une composante saisonnière dans la tendance ! Attention, cependant, la mise à l'échelle automatique des graphiques peut prêter à confusion : à gauche les variations saisonnières vont de -7,5% à +5% ; à droite, il s'agit de $\pm 0,02\%$! Mais les résidus du modèle de droite sont tellement faibles (compris entre $\pm 0,002\%$ que la saisonnalité est considérée comme significative.

On peut remarquer également que les modèles ARIMA retenus sont très différents. En particulier, l'ordre de la partie autorégressive du second modèle est beaucoup plus élevé que celui du premier. Les prévisions fournies par les deux modèles risquent donc de diverger, en particulier si l'horizon de prévision est long.

Cette situation est théoriquement ennuyeuse, pratiquement sans effet. Elle se produit, notamment lorsque des événements ont été détectés et pris en compte dans la série initiale, ce qui est le cas ici. En effet, la « correction des outliers » pratiquée dans Demetra introduit des non linéarités.

Information on Models	Model 1 (Tramo-Seats)	Information on Models	Model 1 (Tramo-Seats)
Time Span (n° of obs.)	Q1.1995 - Q4.2000 (24)	Time Span (n° of obs.)	Q1.1995 - Q4.2000 (24)
Method	Tramo/Seats	Method	Tramo/Seats
PRE-ADJUSTMENT		PRE-ADJUSTMENT	
Transformation	Logarithm	Transformation	Logarithm
Mean Correction	None	Mean Correction	Yes
Correction for Trading Day Effects	None	Mean t-value	2.65 [-2.021, 2.021] 5%
Correction for Easter Effect	None	Correction for Trading Day Effects	None
Correction for Outliers	Autom.:AO,LS,TC; 1 Outlier(...)	Correction for Easter Effect	None
Critical t-value	2.80	Correction for Outliers	Autom.:AO,LS,TC
AO Q4.1995 t-value	3.66 [-2.021, 2.021] 5%	Critical t-value	2.80
Corr. for Missing Obs.	None	Corr. for Missing Obs.	None
Corr. for Other Regr. Effects	None	Corr. for Other Regr. Effects	None
Specif. of the ARIMA model	(0 1 1)(0 1 1) (fixed)	Specif. of the ARIMA model	(3 1 0)(0 1 1) (fixed)
Non-seas. MA (lag 1) value	-0.38	Non-seas. AR (lag 1) value	-1.66
Non-seas. MA (lag 1) t-value	-1.55 [-2.021, 2.021] 5%	Non-seas. AR (lag 1) t-value	-8.73 [-2.021, 2.021] 5%
Seasonal MA (lag 4) value	-0.40	Non-seas. AR (lag 2) value	1.29
Seasonal MA (lag 4) t-value	-1.33 [-2.021, 2.021] 5%	Non-seas. AR (lag 2) t-value	4.95 [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood	Non-seas. AR (lag 3) value	-0.42
DECOMPOSITION		Non-seas. AR (lag 3) t-value	-2.30 [-2.021, 2.021] 5%
ARIMA Decomposition	Exact	Seasonal MA (lag 4) value	-0.69
Seasonality	Seasonal model used	Seasonal MA (lag 4) t-value	-2.24 [-2.021, 2.021] 5%
		Method of Estimation	Exact Maximum Likelihood
		DECOMPOSITION	
		ARIMA Decomposition	Approximated
		Seasonality	Seasonal model used
Information on Diagnostics	Model 1 (Tramo-Seats)	Information on Diagnostics	Model 1 (Tramo-Seats)
Ljung-Box on residuals	5.50 [0, 18.30] 5%	Ljung-Box on residuals	6.06 [0, 15.50] 5%
Box-Pierce on residuals	2.43 [0, 5.99] 5%	Box-Pierce on residuals	1.39 [0, 5.99] 5%
Ljung-Box on squared residuals	9.28 [0, 18.30] 5%	Ljung-Box on squared residuals	9.62 [0, 15.50] 5%
Box-Pierce on squared residuals	0.23 [0, 5.99] 5%	Box-Pierce on squared residuals	1.48 [0, 5.99] 5%
Normality	0.79 [0, 5.99] 5%	Normality	1.42 [0, 5.99] 5%
Skewness	-0.26 [-1.13, 1.13] 5%	Skewness	-0.59 [-1.13, 1.13] 5%
Kurtosis	2.12 [0.74, 5.26] 5%	Kurtosis	3.72 [0.74, 5.26] 5%
Percentage of outliers	4.17% [0%, 5.0%] ad-hoc	Percentage of outliers	0.00% [0%, 5.0%] ad-hoc
Model 1 (Tramo-Seats)		Model 1 (Tramo-Seats)	
-> Model passes all diagnostic tests.		-> Model passes all diagnostic tests.	

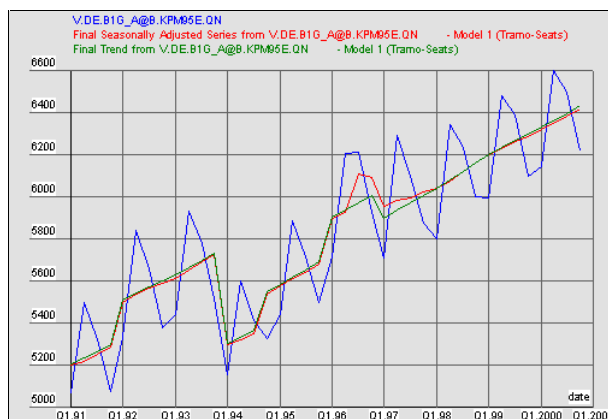
II.6.3. Ljung-Box significatif : structure des résidus



Rupture de tendance, série très longue : raccourcir la série. Il faut donc relancer le traitement en modifiant la période à analyser dans le dialogue de définition d'un nouvel ajustement.

On peut aussi supprimer les points non retenus dans le fichier de données du projet. Ceci permet de conserver une date unique de début pour l'ensemble du projet. En effet, si dans la définition du projet, on spécifie une date de début, celle-ci s'applique à l'ensemble des séries de la table. Bien sûr, Demetra ne traite la série sur la période renseignée.

II.6.4. Trop d'événements (*outliers*)



La série est trop régulière : il faut relever le seuil de détection des points extrêmes, voire supprimer la détection automatique (cf. plus haut, l'amélioration en traitement par lot). Notons, par ailleurs, que nous disposons de dix années de données trimestrielles et que 5% d'*outliers*, valeur pour le rejet d'un ajustement dans les options par défaut sur la qualité des modèles, nous limite à un nombre d'événements strictement inférieur à deux, soit un seul au maximum !



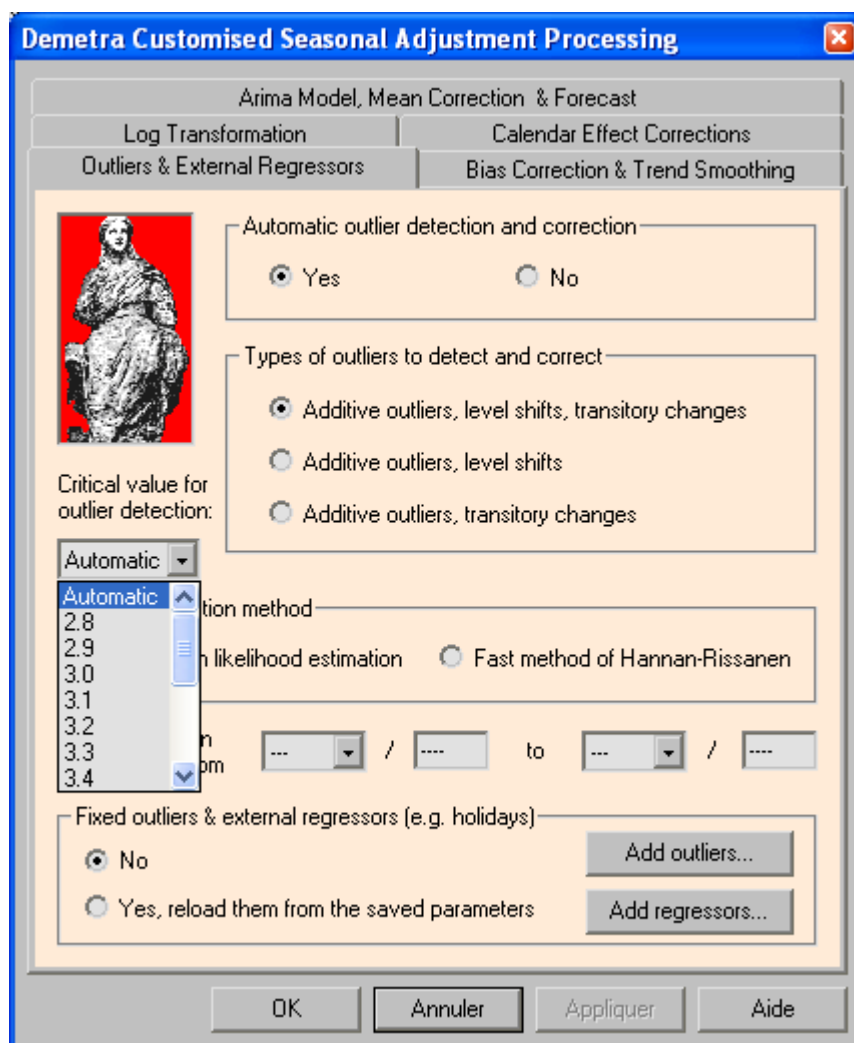
II.6.5. Incorporation d'information *a priori*

Dans le graphique précédent, on peut repérer une dépression de la série sur les deux années 1994 et 1995. Supposons que la cause de cette baisse ait été identifiée, on pourrait alors imaginer « aider » Demetra dans sa recherche d'*outliers* en lui indiquant explicitement la présence de cet événement. Nous allons présenter deux méthodes envisageables :

- 4 introduire deux événements de type LS (sur Q1.94 et Q1.96) ou
- 4 introduire une « variable d'intervention », indiquant qu'il s'est passé quelque chose pendant les deux années.

Notons que ces deux méthodes ne sont pas équivalentes : dans le premier cas, on demande à Demetra de chercher deux échelons, donc deux valeurs de marche d'escalier, *a priori* indépendantes l'une de l'autre. Dans le deuxième, on considère qu'il y a eu un effet sur les deux ans (décalage) et un retour « à la normale » à l'issue de ces deux années ; la valeur de la deuxième marche est l'opposé de celle de la première.

Nous présenterons, enfin, la possibilité d'introduire une variable supplémentaire, choisie ou définie préalablement par l'analyste et dont il suppose qu'elle exerce une influence sur le niveau de la série à modéliser. Nous parlerons alors d'un régresseur externe.

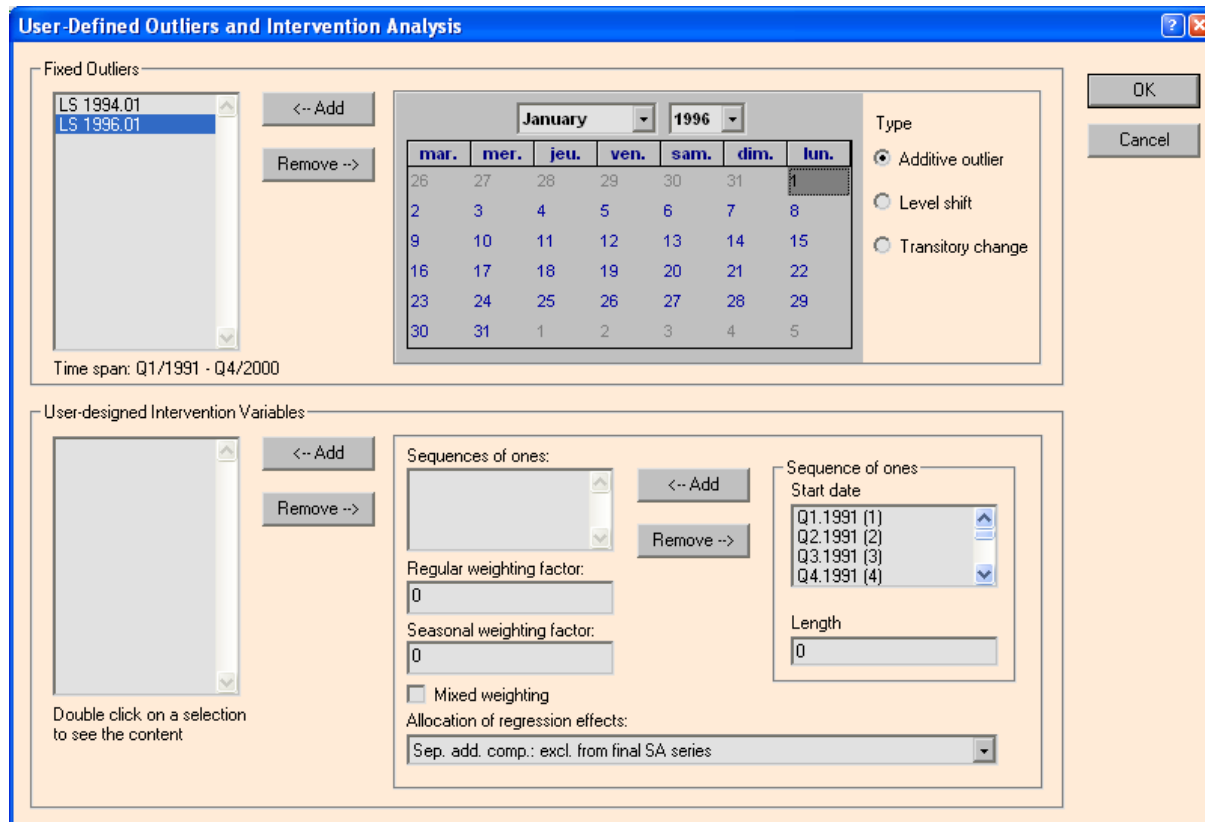




Introduction manuelle d'outliers

Dans le dialogue d'analyse par lot, choisissons l'option *Customised parameters for a new processing*. Dans le dialogue suivant, au bas de l'onglet *Outliers & external regressors*, se trouvent les boutons *Add outliers...* et *Add regressors...*

Le premier nous amène au dialogue suivant.



User-Defined Outliers and Intervention Analysis

Fixed Outliers

LS 1994.01
LS 1996.01

<-- Add
Remove -->

Time span: Q1/1991 - Q4/2000

January 1996

mar.	mer.	jeu.	ven.	sam.	dim.	lun.
26	27	28	29	30	31	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

Type

☒ Additive outlier
☐ Level shift
☐ Transitory change

OK
Cancel

User-designed Intervention Variables

<-- Add
Remove -->

Sequences of ones:

Regular weighting factor: 0
Seasonal weighting factor: 0
☐ Mixed weighting
Allocation of regression effects: Sep. add. comp.: excl. from final SA series

Sequence of ones

Start date

Q1.1991 (1)
Q2.1991 (2)
Q3.1991 (3)
Q4.1991 (4)

Length: 0

Double click on a selection to see the content

La partie haute concerne les événements. Le calendrier permet de choisir les dates des événements connus préalablement par l'analyste. N'oubliez pas de préciser le jour du mois (Demetra se charge de vous le rappeler) et surtout de préciser le type d'événement que vous souhaitez introduire à l'aide des trois boutons radio de droite.

Notez que dans notre exemple, nous introduisons deux événements pour une série dont les données vont de 1991 à 2000, soit 40 trimestres. Nous sommes donc déjà au plafond de 5% d'*outliers* fixé dans les règles de qualité. Il nous faudra donc forcer l'acceptation de ce modèle si les autres critères sont acceptables.

Introduction d'une variable d'intervention

Cette deuxième méthode n'est possible que si l'on utilise l'algorithme Tramo-Seats. X12-Arima ne présentant pas la même souplesse et obligeant à passer par un régresseur défini explicitement. Cette méthode n'est-elle pas accessible à partir du traitement par lot (à la suite probablement d'une bogue (*bug*)). En effet, l'introduction d'une telle variable, représentée par une série de 1, se fait par la partie basse du dialogue représenté ci-dessus. Or, lorsqu'on y accède en provenance du traitement par lot, la zone de liste de droite (*Sequence of ones / Start date*) est vide ou ne contient que la date de début non formatée. Elle devrait contenir la liste des dates observées pour la série. Pour pouvoir tester cette option, il nous faut donc démarrer le module d'amélioration des séries refusées ou basculer la série en mode analyse détaillée. Nous pourrions alors introduire une séquence de 1, débutant en Q1.94 et d'une durée de 8 trimestres.



User-designed Intervention Variables

User intervention 1

<-- Add

Remove -->

Sequences of ones:

1 S:Q1.1994#13 L:8

<-- Add

Remove -->

Regular weighting factor:

0

Seasonal weighting factor:

0

☐ Mixed weighting

Allocation of regression effects:

Assigned to the trend

Sep. add. comp.: excl. from final SA series

Assigned to the trend

Assigned to the seasonal component

Assigned to the irregular component

Sequence of ones

Start date

Q1.1991 (1)

Q2.1991 (2)

Q3.1991 (3)

Q4.1991 (4)

Length

1

Double click on a selection to see the content

N'oublions pas d'indiquer dans la boîte déroulante du bas (*allocation of regression effects*) la composante du modèle qui doit être modifiée par la variable d'intervention dont nous venons de préciser la période d'action. Dans notre cas, nous ferons porter l'intervention sur la tendance.

Tramo-Seats

Separate additional component: exclude from final seasonally adjusted series

Assigned to the trend

Assigned to the seasonal component

Assigned to the irregular component

Assigned to the seas. adjusted series

Assigned to the cycle

Tramo-Seats permet d'attribuer (*assign*) l'effet de la variable à l'une quelconque des composantes du modèle (tendance, saison, irrégulier, cvs ou cycle). En pratique, cela signifie que l'échelon (la marche d'escalier) dont Demetra estimera la hauteur sera considérée comme une correction *a priori* à faire porter sur la composante retenue.

Le premier choix attribue l'effet de l'intervention à une composante particulière qui sera retirée de la série avant toute modélisation et dont l'impact ne figurera pas dans la série cvs.

Notons que vous pouvez vérifier la pertinence de l'introduction de l'information *a priori*. Nous pouvons le constater dans les informations fournies sur les modèles.



Information on Models	Model 3 (Tramo-Seats fixed Outliers)	Model 4 (Tramo-Seats Intervention)
Time Span (n° of obs.)	Q1.1991 - Q4.2000 (40)	Q1.1991 - Q4.2000 (40)
Method	Tramo/Seats	Tramo/Seats
PRE-ADJUSTMENT		
Transformation	None	None
Mean Correction	Yes	None
Mean t-value	4.58 [-2.021, 2.021] 5%	--
Correction for Trading Day Effects	None	None
Correction for Easter Effect	None	None
Correction for Outliers	2 Outlier(s) fixed	None
Critical t-value	3.80	3.20
LS Q1.1994 t-value	-6.89 [-2.021, 2.021] 5%	--
LS Q1.1996 t-value	3.89 [-2.021, 2.021] 5%	--
Corr. for Missing Obs.	None	None
Corr. for Other Regr. Effects	None	1 Regressor(s)
User0 t-value	--	-8.08 [-2.021, 2.021] 5%
Specif. of the ARIMA model	(1 0 0)(0 1 0) (fixed)	(0 1 1)(0 1 0) (fixed)
Non-seas. AR (lag 1) value	-0.55	--
Non-seas. AR (lag 1) t-value	-- [-2.021, 2.021] 5%	--
Non-seas. MA (lag 1) value	--	-0.99
Non-seas. MA (lag 1) t-value	--	-- [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood	Exact Maximum Likelihood
DECOMPOSITION		
ARIMA Decomposition	Exact	Exact
Seasonality	Seasonal model imposed	Seasonal model imposed
Information on Diagnostics	Model 3 (Tramo-Seats fixed Outliers)	Model 4 (Tramo-Seats Intervention)
STATISTICS ON RESIDUALS		
Ljung-Box on residuals	4.44 [0, 25.00] 5%	29.81 [0, 25.00] 5%
Box-Pierce on residuals	0.95 [0, 5.99] 5%	0.72 [0, 5.99] 5%
Ljung-Box on squared residuals	16.74 [0, 25.00] 5%	10.11 [0, 25.00] 5%
Box-Pierce on squared residuals	2.70 [0, 5.99] 5%	0.45 [0, 5.99] 5%
DESCRIPTION OF RESIDUALS		
Normality	2.43 [0, 5.99] 5%	1.34 [0, 5.99] 5%
Skewness	-0.37 [-0.84, 0.84] 5%	-0.18 [-0.82, 0.82] 5%
Kurtosis	4.10 [1.33, 4.67] 5%	2.10 [1.35, 4.65] 5%
OUTLIERS		
Percentage of outliers	5.00% [0%, 5.0%] ad-hoc	0.00% [0%, 5.0%] ad-hoc

L'approche consistant à introduire deux échelons fournit bien deux valeurs significatives (-6,89 et +3,89) différentes en valeur absolue. La deuxième méthode supposant une dépression temporaire de la tendance fournit également une valeur largement significative (-8,08). Attention, ces deux ensembles de valeurs ne sont pas directement comparables entre eux, en effet, les nombres sont exprimés en unités d'écart-type de la composante résiduelle. Or, les deux modèles sont sensiblement différents, les tailles respectives des résidus (leurs écarts-types) le sont donc également. Le choix entre les deux méthodes est essentiellement affaire d'interprétation des événements par l'analyste.

Attention à ne pas mettre en place simultanément les deux méthodes que nous venons d'exposer pour traiter un même événement. Dans la copie d'écran ci-dessous, nous avons imposé *simultanément* les deux événements *et* la modification de tendance.



Information on Models	Model 1 (Tramo-Seats)
Time Span (n° of obs.)	Q1.1991 - Q4.2000 (40)
Method	Tramo/Seats
PRE-ADJUSTMENT	
Transformation	None
Mean Correction	None
Correction for Trading Day Effects	None
Correction for Easter Effect	None
Correction for Outliers	2 Outlier(s) fixed
Critical t-value	2.80
LS Q1.1994 t-value	-195013152.00 [-2.021, 2.0...
LS Q1.1996 t-value	195013152.00 [-2.021, 2.02...
Corr. for Missing Obs.	None
Corr. for Other Regr. Effects	1 Regressor(s)
User0 t-value	195013152.00 [-2.021, 2.02...
Specif. of the ARIMA model	(0 1 1)(0 1 0) (fixed)
Non-seas. MA (lag 1) value	-0.99
Non-seas. MA (lag 1) t-value	-- [-2.021, 2.021] 5%
Method of Estimation	Exact Maximum Likelihood
DECOMPOSITION	
ARIMA Decomposition	Exact
Seasonality	Seasonal model imposed
Information on Diagnostics	Model 1 (Tramo-Seats)
STATISTICS ON RESIDUALS	
Ljung-Box on residuals	18.51 [0, 25.00] 5%
Box-Pierce on residuals	2.67 [0, 5.99] 5%
Ljung-Box on squared residuals	9.29 [0, 25.00] 5%
Box-Pierce on squared residuals	0.79 [0, 5.99] 5%
DESCRIPTION OF RESIDUALS	
Normality	2.53 [0, 5.99] 5%
Skewness	-0.43 [-0.85, 0.85] 5%
Kurtosis	1.92 [1.30, 4.70] 5%
OUTLIERS	
Percentage of outliers	5.00% [0%, 5.0%] ad-hoc

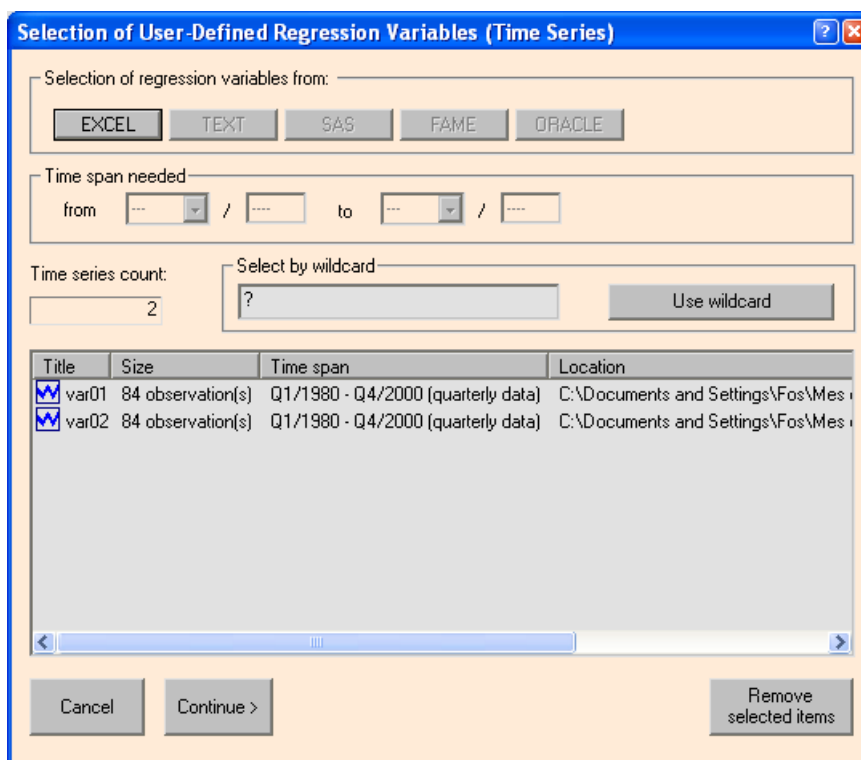
Le résultat signale uniquement le problème du nombre d'événements trop important. Et pourtant, il y a un grave problème : regardez les coefficients des événements et de la variable d'intervention ; ils sont gigantesques ($2 \cdot 10^8$...) et égaux en sens inverse. Il y a, clairement, un problème de colinéarité entre les deux méthodes retenues.

Ø Introduction d'un régresseur externe

La troisième option autorisant l'addition d'information externe connue *a priori* permet d'introduire un régresseur. Il s'agit d'une ou plusieurs variables exerçant une influence sur le niveau de la série à désaisonnaliser. On accède à cette option en cliquant sur le bouton *Add regressors...* de l'onglet *Outliers & external regressors*. L'option est accessible dans les deux grandes familles de méthodes et quel que soit le mode de travail (traitement par lot, amélioration ou analyse détaillée). Nous décrirons son fonctionnement dans le traitement par lot.

Le bouton *Add regressors* appelle un ensemble de deux dialogues : le premier permet de choisir le fichier dans lequel se trouvent la ou les variables externes. Attention, ce fichier est obligatoirement de même type que le fichier qui contient les séries à traiter : seul le bouton correspondant à ce type est activé dans le dialogue.

Sur le modèle du dialogue de sélection des séries à analyser, vous devez indiquer les variables externes que vous allez utiliser pour le modèle de désaisonnalisation. Attention à ne sélectionner que celles qui seront effectivement utilisées.

Selection of User-Defined Regression Variables (Time Series)

Selection of regression variables from:

EXCEL TEXT SAS FAME ORACLE

Time span needed:

from --- / --- to --- / ---

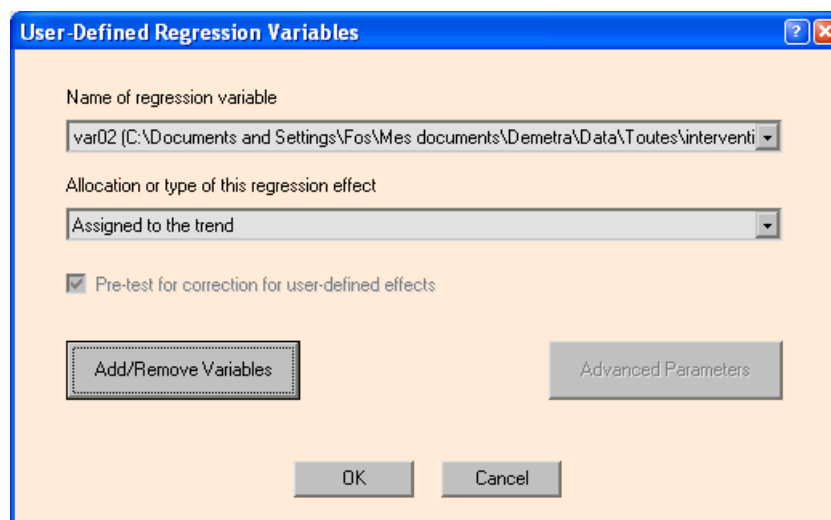
Time series count: 2

Select by wildcard: ? Use wildcard

Title	Size	Time span	Location
var01	84 observation(s)	Q1/1980 - Q4/2000 (quarterly data)	C:\Documents and Settings\Fos\Mes
var02	84 observation(s)	Q1/1980 - Q4/2000 (quarterly data)	C:\Documents and Settings\Fos\Mes

Cancel Continue > Remove selected items

Ayant sélectionné les variables externes, on passe au dialogue suivant par le bouton *Continue*>. On y trouvera, dans la première liste déroulante, la liste des variables sélectionnées dans l'étape précédente.



User-Defined Regression Variables

Name of regression variable

var02 (C:\Documents and Settings\Fos\Mes documents\Demetra\Data\Toutes\interventi)

Allocation or type of this regression effect

Assigned to the trend

☒ Pre-test for correction for user-defined effects

Add/Remove Variables Advanced Parameters

OK Cancel

À chacune d'entre elles, il faudra alors affecter un type d'influence sur la série à désaisonnaliser. Le choix de l'allocation dépend de la méthode employée. Pour Tramo-Seats, on retrouve les options proposées plus haut pour l'introduction de variables d'intervention.



Tramo-Seats

Separate additional component: exclude from final seasonally adjusted series
Assigned to the trend
Assigned to the seasonal component
Assigned to the irregular component
Assigned to the seas. adjusted series
Assigned to the cycle

X12-Arima

User Defined
Constant
User Defined Seasonal
Trading Day
Length-of-Month
Length-of-Quarter
Leap Year
Stock Trading Day
Stock Length-of-Month
Easter
Labor Day
Thanksgiving Day
Additive Outlier
Level Shift
Ramp Effect
User Defined Holiday
Statistics Canada Easter
Temporary Change

Pour X12-Arima, la liste, plus longue, permet de faire porter l'impact de la régression sur différents paramètres du modèle. Les trois premiers éléments de la liste sont équivalents à leurs correspondants disponibles pour la méthode Tramo-Seats. Dans la suite de la liste, on retrouve la possibilité de traiter différents types d'événements (AO, LS, TC et rampe) ou les différents types d'effets de calendrier.

Notons pour terminer que la gestion des *outliers* et régresseurs peut paraître assez lourde : il faut à chaque nouvel essai de modèle réintroduire les événements et les séquences de uns ou sélectionner le fichier contenant les variables. Aussi est-il intéressant d'utiliser l'option (*reload from the saved parameters*) proposée dans le dialogue. Vous pouvez ainsi conserver ces définitions d'informations a priori tout en testant différentes options de modélisation.

Ø Reprise d'un modèle

Lorsque vous traitez un grand nombre de séries par lot dans le module automatisé, il peut se produire que certaines de ces séries soient similaires ou, du moins, fortement reliées entre elles. Ainsi, si vous désaisonnalisiez des statistiques de demandeurs d'emploi, vous êtes probablement amené à travailler à différents niveaux de regroupements des informations : analyse par zone géographique, par sexe, par tranche d'âge. Dans ce contexte, il est fréquent que diverses séries se présentent sous des aspects très voisins.

Or, il est possible que des séries voisines n'aboutissent pas à des modèles de désaisonnalisation similaires, voire que l'un des modèles soit accepté et l'autre refusé. Il peut, alors, être intéressant de sélectionner le modèle retenu pour l'une des séries et de l'appliquer à l'autre. Pour cela trois possibilités s'offrent à vous :

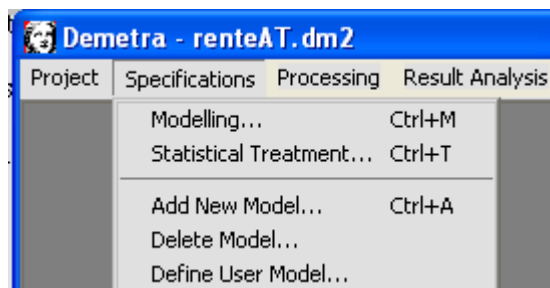
- récupération manuelle de la définition du modèle ou
- par copier-coller dans le fichier de définition du projet
- constitution d'un fichier stockant la définition du modèle (fichier ".mdl").

La première option peut être employée en restant à l'intérieur du module de traitement automatisé : à vous de relever dans les résultats affichés du traitement de la série modèle, les paramètres que vous souhaitez reporter pour l'étude de la deuxième série. Vous choisirez alors un traitement personnalisé de cette dernière et vous parcourrez les onglets du dialogue pour fixer vous même les choix correspondants.

La deuxième option consiste à intervenir dans le fichier de définition du projet (dans Excel, il s'agit de l'onglet *Demetra_parameters* du classeur contenant les données). Repérez la série modèle, vous trouverez la définition, sous forme de texte, de toutes les options spécifiées par l'utilisateur. Copiez ce texte et venez le coller à l'emplacement correspondant pour chacune des séries à paramétrer. Ceci fait, n'oubliez pas dans Demetra de réaliser une mise à jour des séries (dans le tableau des séries du projet faire un clic droit et sélectionner le menu *Time Series->Update...*



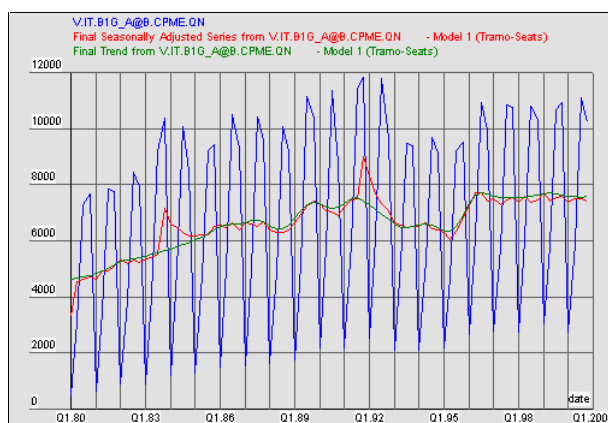
La troisième solution suppose un détournement par le module d'analyse détaillée. Vous allez créer un fichier pour étudier la série modèle. Vous pouvez y mettre au point le modèle d'une manière approfondie. Lorsque vous êtes satisfait du modèle obtenu, vous allez l'exporter en sélectionnant l'option **Define user model...** du menu **Specifications**.



Le dialogue de sélection du modèle dans le traitement détaillé va vous permettre de désigner le modèle à conserver, puis le dialogue d'enregistrement d'un fichier vous permettra de créer un fichier modèle (de type MS-DOS **.mdl**).

Par la suite, vous utiliserez ce modèle en sélectionnant la troisième option du dialogue de choix du traitement (*Parameters from model file for a new processing*) ce qui vous permettra d'aller désigner le fichier à utiliser comme modèle.

II.6.6. Trois statistiques sur les résidus significatives à 5%



Les événements apparents (Q4.83 et Q4.91), tous les deux de type TC, ont l'air surprenants. Le palier des années 1993 à 1995 n'est pas détecté comme événement et il semble y avoir un changement de tendance commençant en 1993.

Les données sont transformées en logarithme ce qui ne semble absolument pas justifié (évolution des données brutes dans un chenal aux bords parallèles).

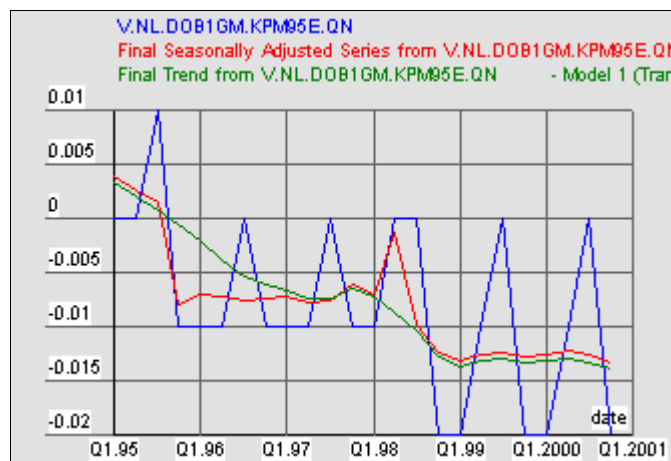
Plusieurs solutions possibles :

- 4 augmenter le seuil de détection des événements,
- 4 ne pas détecter d'événement de type TC, sans changer le seuil,
- 4 bloquer la transformation par les logarithmes.

Chacune de ces actions permet l'acceptation du modèle ainsi obtenu.

II.6.7. Séries non saisonnières

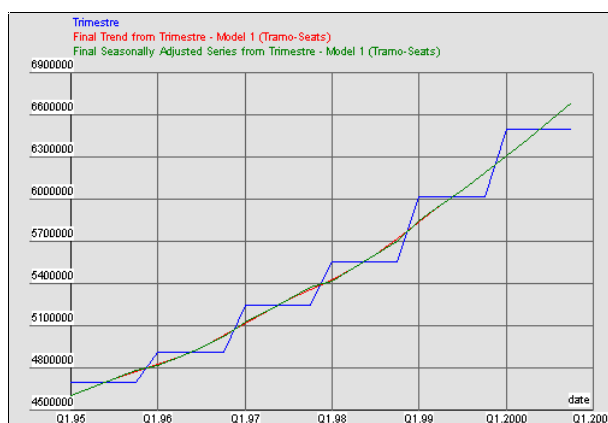
Ø Données discrétisées



Le modèle est rejeté : trop d'*outliers* et une statistique sur les résidus significative. En regardant les données, on peut s'apercevoir qu'il s'agit de valeurs arrondies : sur les 24 valeurs observées, nous n'avons que 4 valeurs distinctes (-2%, -1%, 0% et +1%). Il s'agit probablement de taux de croissance, par exemple de l'indice de prix ou du produit intérieur brut, dont la valeur est publiée en pourcentage, sans décimale. Sur de telles données, discrétisées, il n'est pas prudent d'utiliser des modèles traitant explicitement de variables réelles.

Dans le cas particulier des taux de croissances, on préférera travailler directement sur la variable originale (indice de prix ou du PIB). On peut d'ailleurs remarquer que l'approche de modélisation retenue pour traiter de la série originale intègre la possibilité de l'existence d'un modèle portant sur les variations de la série. C'est la signification du I de ARIMA.

Ø Séries annuelles trimestrialisées



Dans cet exemple, les options par défaut de Tramo-Seats détectent deux statistiques sur les résidus significatives à 5% (en bleu) et un nombre élevé d'*outliers* ce qui entraîne le rejet du modèle. L'amélioration automatique de l'assistant conserve l'excès d'*outliers* et garde d'autres statistiques sur les résidus significatives à 5%.

En optant pour la non détection automatique des *outliers*, ce sont les deux tests de Box-Pierce sur les résidus qui deviennent significatifs, la distribution des résidus est asymétrique ; le modèle est rejeté.



Clairement, il reste une structure dans les résidus ; celle ci est évidente à l'examen de la courbe : les quatre trimestres ont la même valeur. Il s'agit très certainement de données fournies en valeurs annuelles et « trimestrialisées » pour fabriquer une pseudo série en données trimestrielles.

La série n'est pas, en réalité, une série à périodicité d'observation infra annuelle, c'est une série annuelle, il n'y a donc pas lieu de la désaisonnaliser. Il faut donc soit obtenir si elles existent de vraies données, soit supprimer la série du projet.

II.7. Pour conclure

Rappelons ici que le problème de la décomposition saisonnière d'une série chronologique admet une infinité de réponses possibles. Parmi celles-ci, peut-on être sûr d'avoir trouvé le « bon modèle » de désaisonnalisation, voire « le meilleur » ?

Malheureusement, la réponse est non ! Et d'ailleurs, sous quel(s) critère(s) jugerait-on qu'un modèle est « meilleur » qu'un autre ?

Nous avons vu que l'on dispose de nombreux indices qui, s'ils ne désignent pas le bon modèle, permettent parfois de repérer les mauvais ! Nous avons vu également que cette palette d'indices peut, à l'occasion, laisser passer un mauvais modèle. C'est donc ici une dernière occasion de rappeler l'importance extrême de l'examen visuel des séries produites : il serait particulièrement dommage de se priver des immenses capacités d'analyse et de traitement des graphiques dont dispose le cerveau humain.