

IBM SPSS Forecasting 29



Remarque

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations figurant dans la section «Remarques», à la page 53.

Notice d'édition

Cette édition s'applique à la version 29, édition 0, modification 1 d' IBM® SPSS Statistics et à toutes les éditions et modifications ultérieures, sauf indication contraire dans les nouvelles éditions.

© **Copyright International Business Machines Corporation .**

Table des matières

Chapitre 1. Prévisions.....	1
Introduction aux séries chronologiques.....	1
Données de séries temporelles.....	1
Transformations de données.....	2
Périodes d'estimation et de validation.....	2
Création de modèles et prévisions.....	3
Time Series Modeler.....	3
Spécification des options pour le modélisateur expert.....	5
Modèles de lissage exponentiel personnalisés.....	6
Modèles ARIMA personnalisés.....	7
Sortie.....	10
Enregistrement des prévisions et des spécifications de modèle.....	12
Options.....	13
Fonctions supplémentaires de la commande TSMODEL.....	14
Appliquer des modèles de séries temporelles.....	14
Sortie.....	16
Enregistrement des prévisions et des spécifications de modèle.....	19
Options.....	20
Fonctions supplémentaires de la commande TSAPPLY.....	20
Désaisonnalisation.....	20
Enregistrement de la désaisonnalisation.....	21
Fonctions supplémentaires de la commande SEASON.....	22
Tracés spectraux.....	22
Fonctions supplémentaires de la commande SPECTRA.....	24
Modèles de causalité temporelle.....	24
Obtention d'un modèle de causalité temporelle.....	25
Série temporelle à modéliser	25
Observations	27
Intervalle de temps pour l'analyse	28
Agrégation et distribution	28
Valeurs manquantes	29
Options générales pour les données	30
Options générales de construction	30
Séries à afficher	30
Options de sortie	31
Période d'estimation	33
Prévision	34
Enregistrement	34
Sortie interactive	35
Application de modèles de causalité temporelle.....	36
Application de modèles de causalité temporelle.....	36
Prévisions de modèle de causalité temporelle.....	36
Scénarios de modèle de causalité temporelle.....	42
Mesures de la qualité d'ajustement.....	48
Types de valeurs extrêmes.....	49
Guide relatif aux tracés ACF/PACF.....	49
Remarques.....	53
Marques.....	54

Index..... 57

Chapitre 1. Prévisions

Les fonctions de prévision suivantes sont incluses dans SPSS Statistics Professional Edition ou dans l'option Prévisions.

Introduction aux séries chronologiques

Une **série chronologique** est un ensemble d'observations obtenues par mesure régulière d'une variable sur une certaine période. Dans une série de données de stock, par exemple, les observations peuvent représenter des niveaux de stock quotidiens pour plusieurs mois. Une série indiquant la part de marché d'un produit peut représenter la part de marché hebdomadaire relevée sur un certain nombre d'années. Une série comprenant des chiffres relatifs aux ventes totales peut représenter une observation par mois pendant plusieurs années. Tous ces exemples indiquent qu'une variable a été observée à intervalles réguliers au cours d'une certaine période. Par conséquent, les données d'une série chronologique standard représentent une séquence unique ou une liste d'observations représentant des mesures effectuées à intervalles réguliers.

Tableau 1. Série chronologique quotidienne de stock

Durée	Semaine	Jour	Niveau de stock
t ₁	1	Lundi	160
t ₂	1	Mardi	135
t ₃	1	Mercredi	129
t ₄	1	Jeudi	122
t ₅	1	Vendredi	108
t ₆	2	Lundi	150
		...	
t ₆₀	12	Vendredi	120

L'analyse d'une série chronologique est généralement due au besoin de prévoir des valeurs ultérieures. Un modèle de série ayant expliqué les valeurs précédentes peut également indiquer si les valeurs suivantes augmenteront ou diminueront, et à quel degré. La capacité d'effectuer de telles prévisions est évidemment importante pour les secteurs commercial et scientifique.

Données de séries temporelles

Données basées sur les colonnes

Chaque zone de série temporelle contient les données d'une seule série temporelle. Cette structure est la structure traditionnelle des données de série temporelle, telle que celle utilisée par les procédures Time Series Modeler, Désaisonnalisation et Analyse spectrale. Par exemple, pour définir une série chronologique dans l'éditeur de données, cliquez sur l'onglet **Vue de variable**, puis entrez un nom de variable dans une ligne vide. Chaque observation d'une série chronologique correspond à une observation apparaissant sous la forme d'une ligne dans l'éditeur de données.

Si vous ouvrez une feuille de calcul contenant des données de séries temporelles, chaque série doit être disposée dans une colonne de la feuille de calcul. Si vous avez déjà une feuille de calcul contenant des séries chronologiques disposées en lignes, vous pouvez l'ouvrir et utiliser l'option Transposer du menu Données pour transformer les lignes en colonnes.

Données multidimensionnelles

Pour les données multidimensionnelles, chaque champ de série temporelle contient les données de plusieurs séries temporelles. Des séries temporelles distinctes au sein d'un champ donné sont ensuite identifiées par un ensemble de valeurs de champs catégoriels appelés champs de *dimension*.

Par exemple, les données de vente de différentes régions et marques pourraient être stockées dans un champ *ventes* unique, les dimensions dans ce cas correspondant à *région* et à *marque*. Chaque combinaison de *région* et de *marque* identifie une série temporelle spécifique pour *ventes*. Par exemple, dans le tableau suivant, les enregistrements ayant la valeur 'north' comme dimension *région* et la valeur 'brandX' comme dimension *brand* définissent une seule série temporelle.

date	region	marque	ventes
01/01/2014	north	brandX	82350
01/01/2014	north	brandY	86380
01/01/2014	south	brandX	91375
01/01/2014	south	brandY	70320
01/02/2014	north	brandX	83275
01/02/2014	north	brandY	85260
01/02/2014	south	brandX	94760
01/02/2014	south	brandY	69870

Remarque : Les données importées à partir de cubes OLAP, tels que IBM Cognos TM1, sont représentées comme des données multidimensionnelles.

Transformations de données

Un certain nombre de procédures de transformation de données fournies dans le système central sont utiles dans les analyses de séries chronologiques. Ces transformations s'appliquent uniquement aux données basées sur les colonnes où chaque champ de série temporelle contient les données d'une seule série temporelle.

- La procédure Définir des dates (menu Données) génère des variables de date qui sont utilisées pour établir une périodicité et faire la distinction entre les périodes historiques, de validation et de prévision. La prévision est conçue pour être utilisée avec les variables créées par la procédure Définir des dates.
- La procédure Créer la série temporelle (du menu Transformer) crée de nouvelles variables de séries temporelles sous forme de fonctions de variables de séries temporelles existantes. Elle inclut les fonctions qui utilisent les remarques de voisinage pour le lissage, le calcul de la moyenne et la différenciation.
- La procédure Remplacer les valeurs manquantes (du menu Transformer) remplace les valeurs manquantes de l'utilisateur et les valeurs système manquantes par des estimations basées sur une méthode spécifique. Les données manquantes situées au début ou à la fin d'une série ne posent aucun problème : elles réduisent simplement la longueur utile de la série. Les trous situés au milieu d'une série (données manquantes *intégrées*) peuvent représenter un problème beaucoup plus sérieux.

Pour plus d'informations sur la transformation des données des séries temporelles, voir le document *Guide d'utilisation du système central*.

Périodes d'estimation et de validation

Il est souvent utile de diviser une série chronologique en une période *historique* ou *d'estimation* et en une période de *validation*. Développez un modèle sur la base des remarques figurant dans la période d'estimation (historique), puis testez-le pour vérifier son fonctionnement dans la période de validation. En

forçant le modèle à effectuer des prévisions pour les points que vous connaissez déjà (les points de la période de validation), vous obtenez une idée de la capacité de prévision du modèle.

Les observations de la période de validation sont habituellement qualifiées d'observations restantes, car elles sont récupérées du processus de création du modèle. Une fois que la capacité de prévision du modèle vous apporte entière satisfaction, vous pouvez redéfinir la période d'estimation afin d'inclure les observations restantes et de créer ensuite votre modèle final.

Création de modèles et prévisions

Le module complémentaire Prévisions propose les procédures suivantes pour créer des modèles et produire des prévisions :

- La procédure «[Time Series Modeler](#)», à la [page 3](#) crée des modèles pour les séries chronologiques et produit des prévisions. Elle inclut un modélisateur expert qui détermine automatiquement le modèle le plus adapté à chaque série chronologique. Pour les analystes expérimentés, souhaitant bénéficier d'un plus grand degré de contrôle, elle comprend également des outils de génération de modèle personnalisé.
- La procédure décrite au «[Appliquer des modèles de séries temporelles](#)», à la [page 14](#) applique des modèles de séries chronologiques existants, créés à l'aide du Time Series Modeler, à le jeu de données actif. Ceci vous permet d'obtenir des prévisions pour des séries pour lesquelles de nouvelles données ou des données révisées sont disponibles, sans avoir à recréer les modèles. Tout modèle susceptible d'avoir changé peut ainsi être recréé à l'aide du Time Series Modeler.
- La procédure «[Modèles de causalité temporelle](#)», à la [page 24](#) génère des modèles de série temporelle autorégressifs pour chaque cible et détermine automatiquement les meilleures entrées ayant une relation de causalité avec la cible. La procédure crée une sortie interactive que vous pouvez utiliser pour explorer les relations de causalité. La procédure génère également des prévisions, détecte des valeurs extrêmes et détermine les séries les plus susceptibles de causer des valeurs extrêmes.
- La procédure «[Prévisions de modèle de causalité temporelle](#)», à la [page 36](#) applique un modèle de causalité temporelle au jeu de données actif. Cette procédure vous permet d'obtenir des prévisions pour des séries pour lesquelles d'autres données en cours sont disponibles, sans avoir à recréer les modèles. Vous pouvez également l'utiliser pour déterminer les séries les plus susceptibles d'avoir causé les valeurs extrêmes ayant été détectées par la procédure Modèles de causalité temporelle.

Time Series Modeler

La procédure Time Series Modeler évalue le lissage exponentiel, le processus autorégressif moyenne mobile intégré (ARIMA - AutoRegressive Integrated Moving Average) univarié et les modèles ARIMA multivariés (ou modèles des fonctions de transfert) pour les séries chronologiques, et produit des prévisions. La procédure inclut un modélisateur expert qui tente d'identifier et d'évaluer automatiquement le meilleur ARIMA ou modèle de lissage exponentiel pour une ou plusieurs séries de variables dépendantes, éliminant ainsi le besoin d'identifier un modèle approprié via des essais et des erreurs. Vous pouvez également spécifier un modèle ARIMA ou un modèle de lissage exponentiel personnalisé.

Exemple : Vous êtes un chef de produit et êtes responsable des prévisions des ventes de référence et des revenus du mois à venir pour 100 produits distincts et avez peu ou pas du tout d'expérience en modélisation de séries chronologiques. Les données historiques des ventes de référence pour les 100 produits sont stockées dans une feuille de calcul Excel unique. Après avoir ouvert la feuille de calcul dans IBM SPSS Statistics, utilisez le modélisateur expert et demandez les prévisions pour un mois à l'avance. Le modélisateur expert trouve le meilleur modèle de vente de référence pour chacun de vos produits et utilise ces modèles pour produire les prévisions. Le modélisateur expert peut traiter plusieurs séries d'entrées, vous n'avez besoin d'exécuter la procédure qu'une seule fois afin d'obtenir des prévisions pour tous vos produits. En choisissant d'enregistrer les prévisions dans le jeu de données actif, vous pouvez facilement exporter à nouveau les résultats sur Excel.

Statistiques : Mesures de la qualité de l'ajustement : R -carré stationnaire, R -carré (R^2), erreur quadratique moyenne (RMSE), erreur absolue moyenne (MAE), erreur de pourcentage absolue moyenne

(MAPE), erreur absolue maximum (MaxAE), erreur de pourcentage absolue maximum (MaxAPE) et critère d'information bayésien normalisé (BIC). Résidus : fonction d'autocorrélation, fonction d'autocorrélation partielle, Q de Ljung-Box. Pour les modèles ARIMA : ordre ARIMA pour les variables dépendante, ordres de fonction de transfert pour les variables indépendantes et estimations des valeurs extrêmes. De même, le paramètre de lissage estime les modèles de lissage exponentiel.

Tracés : Tracés récapitulatifs parmi tous les modèles : histogrammes de R -carré stationnaire, R -carré (R^2), erreur quadratique moyenne (RMSE), erreur absolue moyenne (MAE), erreur de pourcentage absolue moyenne (MAPE), erreur absolue maximum (MaxAE), erreur de pourcentage absolue maximum (MaxAPE), critère d'information bayésien normalisé (BIC) ; boîtes à moustaches d'autocorrélations résiduelles et partielles. Résultats des modèles individuels : valeurs de prévision, valeurs d'ajustement, valeurs observées, limites de confiance supérieure et inférieure, autocorrélations résiduelles et partielles.

Remarques sur les données du Time Series Modeler

Données : La variable dépendante et les variables indépendantes doivent être numériques.

Hypothèses : La variable dépendante et toute variable indépendante sont traitées en tant que séries chronologiques, ce qui veut dire que chaque observation représente un point dans le temps, avec des observations successives séparées par un intervalle de temps constant.

- **Stationnarité :** Pour les modèles ARIMA personnalisés, les séries chronologiques à modéliser doivent être stationnaires. La méthode la plus efficace pour transformer une série non stationnaire en une série stationnaire nécessite l'utilisation d'une différence, disponible dans la boîte de dialogue Créer la série temporelle .
- **Prévisions :** Pour produire des prévisions à l'aide de modèles avec des variables indépendantes (prédicteur), le jeu de données actif doit contenir des valeurs de ces variables pour toutes les observations de la période de prévision. En outre, les variables indépendantes ne doivent pas contenir de valeurs manquantes dans la période d'estimation.

Définition des dates

Bien que ce ne soit pas une obligation, il est recommandé d'utiliser la boîte de dialogue Définir des dates pour indiquer la date correspondant à la première observation et l'intervalle de temps entre les observations suivantes. Ceci est effectué avant d'utiliser le Time Series Modeler et résulte en un ensemble de variables qui indiquent la date associée à chaque observation. Ceci définit également une périodicité supposée des données, par exemple une périodicité de 12 si l'intervalle de temps entre les observations suivantes est d'un mois. Cette périodicité est requise si vous voulez créer des modèles saisonniers. Si vous ne voulez pas créer de modèles saisonniers et ne voulez pas de libellé de date dans vos sorties, vous pouvez ignorer la boîte de dialogue Définir des dates. Le libellé associé à chaque observation est alors simplement le numéro de l'observation.

Pour utiliser le Time Series Modeler

1. A partir des menus, sélectionnez :

Analyser > Prévision > Créer des modèles traditionnels...

2. Dans l'onglet Variables, sélectionnez une ou plusieurs variables dépendantes à modéliser.
3. Depuis la liste déroulante des méthodes, sélectionnez une méthode de modélisation. Pour la modélisation automatique, conservez la méthode du **modélisateur expert** par défaut. Vous invoquerez ainsi le modélisateur expert pour déterminer le modèle le plus ajusté pour chacune des variables dépendantes.

Pour produire des prévisions :

4. Cliquez sur l'onglet **Options**.
5. Indiquez la période de prévision. Un graphique sera produit, incluant les prévisions et les valeurs observées.

(En option) Vous avez également ces possibilités :

- Sélectionner une ou plusieurs variables indépendantes. Les variables indépendantes sont traitées comme des variables de prédicteur dans les analyses de régression, mais sont facultatives. Elles

peuvent être incluses dans les modèles ARIMA mais pas dans les modèles de lissage exponentiel. Si vous spécifiez le **modélisateur expert** comme méthode de modélisation et incluez des variables indépendantes, seuls les modèles ARIMA seront pris en considération.

- Cliquer sur **Critères** pour indiquer les détails de modélisation.
- Enregistrer les prévisions, les intervalles de confiance et les résidus du bruit.
- Enregistrer les modèles estimés au format XML. Les modèles enregistrés peuvent être appliqués aux données nouvelles ou revus pour obtenir des prévisions mises à jour sans besoin de reconstruire les modèles. .
- Obtenir des statistiques récapitulatives dans tous les modèles estimés.
- Spécifier les fonctions de transfert pour les variables indépendantes dans les modèles ARIMA personnalisés.
- Activer la détection automatique de valeurs extrêmes.
- Points dans le temps spécifiques à un modèle comme valeurs extrêmes pour les modèles ARIMA personnalisés.

Méthodes de modélisation

Les méthodes de modélisation disponibles sont :

Expert Modeler. Le modélisateur expert trouve automatiquement le modèle le plus ajusté pour chaque série dépendante. Si les variables indépendantes (prédicteur) sont indiquées, le modélisateur expert sélectionne des variables ayant une relation statistiquement significative avec la série dépendante pour les inclure dans les modèles ARIMA. Les variables de modèle sont transformées le cas échéant à l'aide de transformation par différenciation, par racine carrée ou par log népérien. Par défaut, le modélisateur expert prend en compte les modèles de lissage exponentiel et les modèles ARIMA. Toutefois, vous pouvez limiter le modélisateur expert pour qu'il ne recherche que les modèles ARIMA ou les modèles de lissage exponentiel. Vous pouvez également indiquer la détection automatique des valeurs extrêmes.

Lissage exponentiel. Utilisez cette option pour indiquer un modèle de lissage exponentiel personnalisé. Vous pouvez choisir parmi une variété de modèles exponentiels qui diffèrent dans leur manière de traiter la tendance et la saison.

ARIMA. Utilisez cette option pour indiquer un modèle ARIMA personnalisé. Ceci implique de spécifier de manière explicite les ordres autorégressifs et de moyenne mobile ainsi que le degré de différenciation. Vous pouvez inclure des variables indépendantes (prédicteur) et définir les fonctions de transfert pour certaines d'entre elles ou pour la totalité. Vous pouvez également spécifier la détection automatique des valeurs extrêmes ou spécifier un ensemble explicite de valeurs extrêmes.

Périodes d'estimation et de prévision

Période d'estimation : La période d'estimation définit l'ensemble des observations utilisées pour déterminer le modèle. Par défaut la période d'estimation inclut toutes les observations dans le jeu de données actif. Pour définir la période d'estimation, sélectionnez **Dans une plage de temps ou d'observations** dans la boîte de dialogue Sélectionner des observations. En fonction des données disponibles, la période d'estimation utilisée par la procédure peut varier selon la variable dépendante et donc être différente de la valeur affichée. Pour une variable dépendante donnée, la vraie période d'estimation est la période restante après l'élimination de toutes les valeurs manquantes contiguës de la variable qui se produit au début ou à la fin de la période d'estimation spécifiée.

Période de prévision : La période de prévision commence à la première observation après chaque période d'estimation et, par défaut, continue jusqu'à la dernière observation dans le jeu de données actif. Vous pouvez définir la fin d'une période de prévision à partir de l'onglet Options.

Spécification des options pour le modélisateur expert

Le modélisateur expert fournit des options pour contraindre l'ensemble de modèles possibles, spécifiant le traitement des valeurs extrêmes, y compris les variables d'événement.

Sélection de modèle et spécification d'événement

L'onglet **Modèle** vous permet d'indiquer les types de modèles pris en compte par le modélisateur expert et de spécifier les variables d'événement.

Type de modèle. Les options suivantes sont disponibles :

- **Tous les modèles.** Le modélisateur expert considère aussi bien les modèles ARIMA que ceux de lissage exponentiel.
- **Modèles de lissage exponentiel uniquement :** Le modélisateur expert considère uniquement les modèles de lissage exponentiel.
- **Modèles ARIMA uniquement.** Le modélisateur expert considère uniquement les modèles ARIMA.

Expert Modeler prend en compte les modèles saisonniers. Cette option est uniquement activée si une périodicité a été définie pour l'ensemble de données actif. Une fois cette option sélectionnée (cochée), le modélisateur expert considère les modèles à la fois saisonniers et non saisonniers. Si cette option n'est pas sélectionnée, Expert Modeler considère uniquement les modèles non saisonniers.

Périodicité actuelle. Indique la périodicité courante (le cas échéant) définie pour le jeu de données actif. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité à partir de la boîte de dialogue Définir des dates.

Événements : Sélectionnez les variables indépendantes qui doivent être traitées comme variables d'événement. Pour les variables d'événement, les observations avec une valeur de 1 indiquent les moments auxquels les séries indépendantes seront affectées par l'événement. Les valeurs autres que 1 n'indiquent aucun effet.

Traitement des valeurs extrêmes avec le modélisateur expert

L'onglet **Valeurs extrêmes** vous permet de choisir la détection automatique de valeurs extrêmes ainsi que le type de valeurs extrêmes à détecter.

Détecter les valeurs extrêmes automatiquement : Par défaut, la détection automatique de valeurs extrêmes n'est pas exécutée. Sélectionnez (cochez) cette option pour exécuter la détection automatique de valeurs extrêmes et sélectionnez un ou plusieurs des types de valeurs extrêmes suivants :

- Additive
- Changement de niveau
- Innovation
- Transitoire
- Additif saisonnier
- Tendance locale
- Additive corrective

Modèles de lissage exponentiel personnalisés

Type de modèle. modèle de lissage exponentiel¹ sont classés comme saisonniers ou non saisonniers. Les modèles saisonniers sont uniquement disponibles si une périodicité a été définie pour le jeu de données actif (voir « Périodicité courante » ci-dessous).

- *Simple.* Ce modèle est approprié pour les séries ne contenant aucune tendance ni saison. Son seul paramètre de lissage est le niveau. Le lissage exponentiel simple est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, un ordre de différenciation, un ordre de moyenne mobile et aucune constante.

¹ Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1-28.

- *Tendance linéaire de Holt*. Ce modèle est approprié aux séries dont se dégage une tendance linéaire, mais aucun effet saisonnier. Ses paramètres de lissage sont le niveau et la tendance qui ne sont pas contraints par leurs valeurs réciproques. Le modèle de Holt est plus général que le modèle de Brown, mais prend plus de temps à calculer pour les grandes séries. Le lissage exponentiel de Holt est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, deux ordres de différenciation et deux ordres de moyenne mobile.
- *tendance linéaire de Brown*. Ce modèle est approprié aux séries dont se dégage une tendance linéaire, mais aucun effet saisonnier. Ses paramètres de lissage sont le niveau et la tendance que l'on considère comme égaux. Par conséquent, le modèle de Brown est un cas particulier du modèle de Holt. Le lissage exponentiel de Brown est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, deux ordres de différenciation et deux ordres de moyenne mobile, où le coefficient pour le second ordre de moyenne mobile est égal au carré de la moitié du coefficient du premier ordre.
- *tendance réduite*. Ce modèle est approprié aux séries présentant une tendance linéaire qui disparaît peu à peu et qui est dépourvue d'effets saisonniers. Ses paramètres de lissage sont le niveau, la tendance et l'oscillation amortie. Le modèle de lissage exponentiel est très similaire à un modèle ARIMA avec un ordre d'autorégression, 1 ordre de différenciation et 2 ordres de moyenne mobile.
- *saisonnier simple*. Ce modèle est approprié pour les séries sans tendance et avec un effet saisonnier constant dans le temps. Ses paramètres de lissage sont le niveau et la saison. Le lissage exponentiel saisonnier simple est similaire à un modèle ARIMA d'ordre 0 pour le terme d'auto régression, d'ordre 1 pour la différenciation première et d'ordre 1, p et p+1 pour la moyenne mobile, où p est le nombre de périodes dans l'intervalle saisonnier (pour des données mensuelles, p = 12).
- *additif de Winters*. Ce modèle est approprié pour les séries avec une tendance linéaire et un effet saisonnier qui ne dépend pas du niveau de la série. Ses paramètres de lissage sont le niveau, la tendance et la saison. Le lissage exponentiel additif de Winters est similaire à un modèle ARIMA d'ordre 0 pour le terme d'auto régression, d'ordre 1 pour la différenciation première, d'ordre 1 pour la différenciation saisonnière, et p+1 pour la moyenne mobile, où p est le nombre de périodes dans l'intervalle saisonnier (pour des données mensuelles, p = 12).
- *mutiplicatif de Winters*. Ce modèle est approprié pour les séries avec une tendance linéaire et un effet saisonnier qui dépend du niveau de la série. Ses paramètres de lissage sont le niveau, la tendance et la saison. Le lissage exponentiel multiplicatif de Winters n'est pas similaire à un modèle ARIMA.

Périodicité actuelle. Indique la périodicité courante (le cas échéant) définie pour le jeu de données actif. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité à partir de la boîte de dialogue Définir des dates.

Transformation de variables dépendantes : Vous pouvez spécifier une transformation exécutée sur chaque variable dépendante avant de la modéliser.

- **Aucun :** Aucune transformation n'est effectuée.
- **Racine carrée.** Transformation racine carrée.
- **Log naturel.** Transformation par log naturel.

Modèles ARIMA personnalisés

Le modélisateur Time Series vous permet de générer des modèles ARIMA (Autoregressive Integrated Moving Average) non saisonniers ou saisonniers, également connus sous le nom de Box-Jenkins² Modèles--avec ou sans un ensemble fixe de variables de prédicteur. Vous pouvez définir des fonctions de transfert pour une ou toutes les variables de prédicteur et indiquer la détection automatique de valeurs extrêmes ou spécifier un ensemble explicite de valeurs extrêmes.

- Toutes les variables indépendantes (prédicteur) indiquées dans l'onglet Variables sont explicitement incluses dans le modèle. Ceci se démarque de l'utilisation du modélisateur expert dans lequel les

² Box, G. E. P., G. M. Jenkins et G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3e éd. Englewood Cliffs, N.J.: Prentice Hall.

variables indépendantes sont uniquement incluses si elles ont une relation statistiquement significative avec la variable dépendante.

Spécification de modèle pour les modèles ARIMA personnalisés

L'onglet Modèle vous permet de spécifier la structure d'un modèle ARIMA personnalisé.

Ordres ARIMA. Entrez des valeurs pour les différentes composantes ARIMA de votre modèle dans les cellules correspondantes de la grille Structure. Toutes les valeurs doivent être des entiers non négatifs. Pour les composants autorégressifs et de moyenne mobile, la valeur représente l'ordre maximum. Tous les ordres inférieurs positifs seront inclus dans le modèle. Par exemple, si vous indiquez 2, le modèle inclut les commandes 2 et 1. Les cellules de la colonne Saisonnier ne sont activées que si une périodicité a été définie pour le jeu de données actif (voir « Périodicité courante » ci-dessous).

- **Autorégressif (p).** Le nombre d'ordres autorégressifs dans le modèle. Les ordres autorégressifs indiquent quelles valeurs précédentes de la série seront utilisées pour prévoir les valeurs en cours. Par exemple, un ordre autorégressif de 2 indique que la valeur de la série Deux points dans le temps dans le passé sera utilisée pour prévoir la valeur en cours.
- **Différence (d) :** Spécifie l'ordre de différenciation appliqué à la série avant d'estimer les modèles. La différenciation est nécessaire lorsque les tendances sont présentes (les séries avec tendances sont en général non stationnaires et la modélisation ARIMA suppose la stationnarité) et est utilisée pour supprimer leurs effets. L'ordre de différenciation correspond au degré de tendance de série, aux comptes de différenciation de premier ordre pour les tendances linéaires, aux comptes de différenciation de second ordre pour les tendances quadratiques, etc.
- **Moyenne mobile (q) :** Le nombre d'ordres de moyenne mobile dans le modèle. Les ordres de moyenne mobile indiquent comment les déviations de la moyenne de la série pour les valeurs précédentes sont utilisés pour prévoir les valeurs courantes. Par exemple, des ordres de moyenne mobile de 1 et 2 indiquent que les écarts de la valeur de la moyenne de la série pour chacune des deux dernières périodes doivent être considérés lors de la prévision des valeurs actuelles de la série.

Ordres saisonniers : Les composants autorégressifs, de moyenne mobile et de différenciation saisonniers tiennent le même rôle que leurs équivalents non saisonniers. Cependant, pour les ordres saisonniers, les valeurs courantes de la série sont affectées par les valeurs de série précédentes séparées par une ou plusieurs périodes saisonnières. Par exemple, pour des données mensuelles (période saisonnière de 12), un ordre saisonnier de 1 indique que la valeur de série en cours est affectée par les 12 périodes de la valeur de série précédant celle en cours. Un ordre saisonnier de 1, pour des données mensuelles, est alors le même que lorsqu'on spécifie un ordre non saisonnier de 12.

Périodicité actuelle. Indique la périodicité courante (le cas échéant) définie pour le jeu de données actif. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité à partir de la boîte de dialogue Définir des dates.

Transformation de variables dépendantes : Vous pouvez spécifier une transformation exécutée sur chaque variable dépendante avant de la modéliser.

- **Aucun :** Aucune transformation n'est effectuée.
- **Racine carrée.** Transformation racine carrée.
- **Log naturel.** Transformation par log naturel.

Inclure la constante dans le modèle. L'inclusion d'une constante est standard, sauf si vous êtes sûr que la valeur moyenne globale de la série est 0. L'exclusion de la constante est recommandée lors de l'application de la différenciation.

Fonctions de transfert dans des modèles ARIMA personnalisés

L'onglet Fonction de transfert (uniquement présent si les variables indépendantes sont spécifiées) vous permet de définir les fonctions de transfert pour une ou toutes les variables indépendantes spécifiées dans l'onglet Variables. Les fonctions de transfert vous permettent de spécifier la manière avec laquelle

les valeurs passées des variables indépendantes (prédicteur) sont utilisées pour prévoir les valeurs futures de la série dépendante.

Ordres de fonction de transfert : Entrez des valeurs pour les différents composants de la fonction de transfert dans les cellules correspondantes de la grille Structure. Toutes les valeurs doivent être des entiers non négatifs. Pour les composants numérateur et dénominateur, la valeur représente l'ordre maximum. Tous les ordres inférieurs positifs seront inclus dans le modèle. En outre, l'ordre 0 est toujours inclus pour les composants numérateur. Par exemple, si vous indiquez 2 pour le numérateur, le modèle inclut les ordres 2, 1 et 0. Si vous indiquez 3 pour le dénominateur, le modèle inclut les ordres 3, 2 et 1. Les cellules de la colonne Saisonnier ne sont activées que si une périodicité a été définie pour le jeu de données actif (voir « Périodicité courante » ci-dessous).

- **Numérateur :** L'ordre de numérateur de la fonction de transfert. Spécifie quelles valeurs précédentes de la série indépendante (prédicteur) sélectionnée sont utilisées pour prévoir les valeurs courantes de la série dépendante. Par exemple, un ordre de numérateur de 1 indique que la valeur d'une série indépendante Un point dans le temps dans le passé, ainsi que la valeur courante de la série indépendante, sont utilisées pour prévoir la valeur courante de chaque série dépendante.
- **Dénominateur :** L'ordre de dénominateur de la fonction de transfert. Spécifie comment les déviations de la moyenne de la série, pour les valeurs précédentes de la série indépendante (prédicteur) sélectionnée, sont utilisés pour prévoir les valeurs courantes de la série dépendante. Par exemple, un ordre de dénominateur de 1 indique que les écarts de la valeur moyenne d'une série indépendante Un point dans le temps dans le passé doivent être considérés lors de la prévision de la valeur en cours de chaque série dépendante.
- **Différence :** Spécifie l'ordre de différenciation appliqué à la série indépendante (prédicteur) avant d'estimer les modèles. La différenciation est nécessaire lorsque les tendances sont présentes et est utilisée pour supprimer leur effet.

Ordres saisonniers : Les composants numérateur, dénominateur et de différenciation saisonniers tiennent le même rôle que leurs équivalents non saisonniers. Cependant, pour les ordres saisonniers, les valeurs courantes de la série sont affectées par les valeurs de série précédentes séparées par une ou plusieurs périodes saisonnières. Par exemple, pour des données mensuelles (période saisonnière de 12), un ordre saisonnier de 1 indique que la valeur de série en cours est affectée par les 12 périodes de la valeur de série précédant celle en cours. Un ordre saisonnier de 1, pour des données mensuelles, est alors le même que lorsqu'on spécifie un ordre non saisonnier de 12.

Périodicité actuelle. Indique la périodicité courante (le cas échéant) définie pour le jeu de données actif. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité à partir de la boîte de dialogue Définir des dates.

Délai : La définition d'un délai retarde l'influence de la variable indépendante selon le nombre d'intervalles spécifié. Par exemple, si le délai est défini sur 5, la valeur de la variable indépendante au moment t n'affecte pas les prévisions tant que les cinq périodes ne se sont pas écoulées ($t + 5$).

Transformation : La spécification d'une fonction de transfert, pour un ensemble de variables indépendantes, inclut également l'exécution d'une transformation optionnelle sur ces variables.

- **Aucun :** Aucune transformation n'est effectuée.
- **Racine carrée.** Transformation racine carrée.
- **Log naturel.** Transformation par log naturel.

Valeurs extrêmes dans les modèles ARIMA personnalisés

L'onglet Valeurs extrêmes fournit les choix suivants pour le traitement des valeurs éloignées³: les détecter automatiquement, spécifier des points particuliers comme valeurs aberrantes, ou ne pas les détecter ou les modéliser.

³ Pena, D., G. C. Tiao, et R. S. Tsay, eds. 2001. *A course in time series analysis*. New York : John Wiley and Sons.

Ne pas détecter les valeurs extrêmes ou les modéliser. Par défaut, les valeurs extrêmes ne sont ni détectées ni modélisées. Sélectionnez cette option pour désactiver la détection ou la modélisation des valeurs extrêmes.

Détecter les valeurs extrêmes automatiquement : Sélectionnez cette option pour exécuter la détection automatique des valeurs extrêmes et sélectionnez un ou plusieurs des types de valeurs extrêmes suivants :

- Additive
- Changement de niveau
- Innovation
- Transitoire
- Additif saisonnier
- Tendance locale
- Additive corrective

Points dans le temps spécifiques à un modèle comme valeurs extrêmes : Sélectionnez cette option pour indiquer des points dans le temps particuliers comme valeurs extrêmes. Utilisez une ligne distincte de la grille Définition des valeurs extrêmes pour chaque valeur extrême. Entrez des valeurs dans toutes les cellules d'une ligne donnée.

- **Type.** Le type de valeur extrême. Les types pris en charge sont : additif (valeur par défaut), changement de niveau, innovation, transitoire, additive saisonnière et tendance locale.

Remarque 1 : Si aucune spécification de date n'a été définie pour le jeu de données actif, la grille Définition des valeurs extrêmes affiche la colonne unique *Observation*. Pour indiquer une valeur extrême, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

Remarque 2 : La colonne *Cycle* (si elle présente) dans la grille Définition des valeurs extrêmes se rapporte à la valeur de la variable *CYCLE_* dans le jeu de données actif.

Sortie

Les sorties disponibles incluent des résultats pour les modèles individuels ainsi que des résultats calculés parmi tous les modèles. Les résultats pour les modèles individuels peuvent être limités à l'ensemble des modèles les plus ou les moins ajustés en fonction de critères définis par l'utilisateur.

Tableaux de statistiques et de prévision

L'onglet Statistiques fournit des options d'affichage pour les tableaux des résultats de la modélisation.

Afficher les mesures d'ajustement, la statistique Ljung-Box et le nombre de valeurs extrêmes par modèle : Sélectionnez (cochez) cette option pour afficher un tableau contenant les mesures d'ajustement, la valeur Ljung-Box et le nombre de valeurs extrêmes sélectionnées pour chaque modèle estimé.

Mesures d'ajustement : Vous pouvez sélectionner une ou plusieurs des valeurs suivantes pour les inclure dans le tableau contenant les mesures d'ajustement pour chaque modèle estimé :

- *R*-carré stationnaire
- *R*-carré
- Erreur quadratique moyenne
- Erreur de pourcentage absolue moyenne
- Erreur Absolue Moyenne
- Erreur de pourcentage absolue maximum
- Erreur Absolue Maximum
- BIC normalisé

Statistiques de comparaison des modèles : Ce groupe d'options contrôle l'affichage des tableaux contenant des statistiques calculées parmi tous les modèles estimés. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Qualité de l'ajustement.** Tableau des statistiques récapitulatives et des percentiles pour R -square fixe, R -square, erreur quadratique moyenne quadratique, erreur absolue moyenne en pourcentage, erreur absolue moyenne, erreur absolue maximale en pourcentage, erreur absolue maximale et critère d'information bayésien normalisé.
- **Fonction d'autocorrélation (ACF) résiduelle :** Tableau des statistiques et percentiles récapitulatifs pour les autocorrélations résiduelles dans tous les modèles estimés.
- **Fonction d'autocorrélation partielle (PACF) résiduelle :** Tableau des statistiques et percentiles récapitulatifs pour les autocorrélations partielles résiduelles dans tous les modèles estimés.

Statistiques de modèles individuels : Ce groupe d'options contrôle l'affichage des tableaux contenant des informations détaillées pour chaque modèle estimé. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Estimations des paramètres.** Affiche un tableau d'estimations des paramètres pour chaque modèle estimé. Des tableaux distincts sont affichés pour les modèles de lissage exponentiel et ARIMA. Si des valeurs extrêmes existent, leurs estimations de paramètres sont également affichées dans un tableau distinct.
- **Fonction d'autocorrélation (ACF) résiduelle :** Affiche un tableau des autocorrélations résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations.
- **Fonction d'autocorrélation partielle (PACF) résiduelle :** Affiche un tableau des autocorrélations partielles résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations partielles.

Afficher les prévisions : Affiche un tableau des prévisions du modèle et les intervalles de confiance pour chaque modèle estimé. La période de prévision est définie dans l'onglet Options.

Tracés

L'onglet Tracés fournit des options d'affichage pour les tracés des résultats de la modélisation.

Tracés de comparaison des modèles

Ce groupe d'options contrôle l'affichage des tracés contenant des statistiques calculées parmi tous les modèles estimés. Chaque option génère un tracé distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- R -carré stationnaire
- R -carré
- Erreur quadratique moyenne.
- Erreur absolue moyenne en pourcentage
- Erreur absolue moyenne.
- Erreur de pourcentage maximale
- Erreur absolue Maximum
- BIC normalisé
- Fonction d'autocorrélation (ACF) résiduelle
- Fonction d'autocorrélation partielle (PACF) résiduelle

Tracés de modèles individuels

Série : Sélectionnez (cochez) cette option pour obtenir des tracés des prévisions pour chaque modèle estimé. Vous pouvez sélectionner un ou plusieurs des éléments suivants pour les insérer dans le tracé :

- **Valeurs observées :** Les valeurs observées des séries dépendantes.

- **Prévisions** : Les prévisions du modèle pour la période de prévision.
- **Valeurs d'ajustement** : Les prévisions du modèle pour la période d'estimation.
- **Intervalle de confiance pour les prévisions** : Les intervalles de confiance pour la période de prévision.
- **Intervalle de confiance pour les valeurs d'ajustement** : Les intervalles de confiance pour la période d'estimation.

Fonction d'autocorrélation (ACF) résiduelle : Affiche un tracé des autocorrélations résiduelles pour chaque modèle estimé.

Fonction d'autocorrélation partielle (PACF) résiduelle : Affiche un tracé des autocorrélations partielles résiduelles pour chaque modèle estimé.

Limitation des sorties aux modèles les plus ou les moins ajustés

L'onglet Filtre des sorties fournit des options pour restreindre à la fois des sorties de tableau ou de graphique à un sous-ensemble de modèles estimés. Vous pouvez choisir de limiter des sorties aux modèles les plus ou les moins ajustés en fonction de critères d'ajustement que vous fournissez. Par défaut, tous les modèles estimés sont inclus dans les sorties.

Modèles les plus appropriés : Sélectionnez (cochez) cette option pour inclure les modèles les plus ajustés dans les sorties. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les moins ajustés. Dans ce cas, les sorties concernent aussi bien les modèles les moins ajustés que les modèles les plus ajustés.

- **Nombre fixe de modèles** : Indique que les résultats sont affichés pour les n modèles les plus ajustés. Si le nombre dépasse le nombre de modèles estimés, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles** : Indique que les résultats sont affichés pour les modèles possédant des valeurs de qualité d'ajustement comprises dans les n pourcentages supérieurs dans tous les modèles estimés.

Modèles les moins ajustés : Sélectionnez (cochez) cette option pour inclure les modèles les moins ajustés dans les sorties. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les plus ajustés. Dans ce cas, les sorties sont constituées des modèles les plus ajustés ainsi que des modèles les moins ajustés.

- **Nombre fixe de modèles** : Indique que les résultats sont affichés pour les n modèles les moins ajustés. Si le nombre dépasse le nombre de modèles estimés, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles** : Indique que les résultats sont affichés pour les modèles possédant des valeurs de qualité d'ajustement comprises dans les n pourcentages inférieurs dans tous les modèles estimés.

Mesure de la qualité de l'ajustement : Sélectionnez la mesure de la qualité d'ajustement à utiliser pour filtrer les modèles. La valeur par défaut est R -carré stationnaire.

Enregistrement des prévisions et des spécifications de modèle

L'onglet Enregistrer vous permet d'enregistrer des prévisions de modèle en tant que nouvelles variables dans le jeu de données actif et d'enregistrer les spécifications de modèle dans un fichier externe au format XML.

Enregistrer les variables : Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résidus en tant que nouvelles variables dans le jeu de données actif. Chaque série dépendante génère son propre ensemble de nouvelles variables et chaque nouvelle variable contient des valeurs pour les périodes de prévision et d'estimation. Les nouvelles observations sont ajoutées si la période de prévision s'étend au-delà de la longueur des séries de variables dépendantes. Choisissez d'enregistrer de nouvelles variables en cochant la case Enregistrer correspondant à chaque variable. Par défaut, aucune nouvelle variable n'est enregistrée.

- **Prévisions** : Les prévisions de modèle.
- **Limites de confiance inférieures** : Limites de confiance inférieures pour les prévisions.
- **Limites de confiance supérieures** : Limites de confiance supérieures pour les prévisions.
- **Résidus du bruit** : Les résidus de modèle. Lors de la transformation des variables dépendantes (par exemple Log népérien), ces derniers sont les résidus des séries transformées.
- **Préfixe de nom de variable** : Indiquez les préfixes à utiliser pour des nouveaux noms de variable ou conservez les préfixes par défaut. Les noms de variable sont constitués du préfixe, du nom de la variable dépendante correspondante et d'un identificateur de modèle. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables. Le préfixe doit se conformer aux règles pour les noms de variable valides.

Exporter un fichier de modèle : Les spécifications de modèle pour tous les modèles estimés sont exportées vers le fichier spécifié au format XML. Les modèles enregistrés peuvent être utilisés pour obtenir des prévisions mises à jour.

- **Fichier XML** : Les spécifications de modèle sont enregistrées dans un fichier XML qui peut être utilisé avec les applications IBM SPSS.
- **Fichier PMML** : Les spécifications de modèle sont enregistrées dans un fichier XML compatible-PMML qui peut être utilisé avec les applications compatibles PMML, y compris les applications IBM SPSS.

Options

L'onglet Options vous permet de définir la période de prévision, spécifier le traitement des valeurs manquantes, définir la largeur de l'intervalle de confiance, spécifier un préfixe personnalisé pour les identificateurs de modèle et définir le nombre de décalages affiché pour les autocorrélations.

Période de prévision : La période de prévision commence toujours par la première observation après la fin de la période d'estimation (l'ensemble d'observations utilisé pour déterminer le modèle) et continue jusqu'à la dernière observation du jeu de données actif ou jusqu'à une date indiquée par l'utilisateur. Par défaut, la fin de la période d'estimation est la dernière observation dans le jeu de données actif, mais peut être modifiée via la boîte de dialogue Sélectionner des observations en sélectionnant **Dans une plage de temps ou d'observations**.

- **Première observation après la fin de la période d'estimation jusqu'à la dernière observation du jeu de données actif** : Sélectionnez cette option lorsque la fin de la période d'estimation est antérieure à la dernière observation du jeu de données actif et que vous souhaitez avoir des prévisions jusqu'à la dernière observation. Cette option est généralement utilisée pour produire des prévisions pour une période restante, permettant une comparaison des prévisions du modèle avec un sous-ensemble des valeurs courantes.
- **Première observation après la fin de la période d'estimation jusqu'à une date spécifique** : Sélectionnez cette option pour indiquer explicitement la fin de la période de prévision. Cette option est généralement utilisée pour produire des prévisions après la fin de la série en cours. Entrez les valeurs dans toutes les cellules de la grille Date.

Si aucune spécification de date n'a été définie pour le jeu de données actif, la grille Date affiche la colonne unique *Observation*. Pour indiquer la fin de la période de prévision, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

La colonne *Cycle* (si elle est présente) dans la grille Date se rapporte à la valeur de la variable *CYCLE_* du jeu de données actif.

Valeurs manquantes de l'utilisateur : Ces options contrôlent le traitement des valeurs manquantes de l'utilisateur.

- **Traiter comme non valide** : Les valeurs manquantes de l'utilisateur sont traitées comme des valeurs système manquantes.
- **Traiter comme valide** : Les valeurs manquantes de l'utilisateur sont traitées comme des données valides.

Stratégie de valeur manquante : Les règles suivantes s'appliquent au traitement des valeurs manquantes (inclut les valeurs système manquantes et les valeurs manquantes de l'utilisateur traitées comme non valides) lors de la procédure de modélisation :

- Les observations avec des valeurs manquantes de variable dépendante qui se produisent au cours de la période d'estimation sont incluses dans le modèle. Le traitement spécifique de la valeur manquante dépend de la méthode d'estimation.
- Un avertissement est émis si une variable indépendante a des valeurs manquantes pour la période d'estimation. Pour le modélisateur expert, les modèles utilisant la variable indépendante sont estimés sans la variable. Pour ARIMA personnalisé, les modèles utilisant la variable indépendante ne sont pas estimés.
- Si une variable indépendante a des valeurs manquantes pour la période d'estimation, la procédure émet un avertissement et effectue les prévisions dans la limite du possible.

Largeur de l'intervalle de confiance (%). Les intervalles de confiance sont calculés pour les prédictions du modèle et les autocorrélations résiduelles. Vous pouvez spécifier n'importe quelle valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Préfixe pour les identificateurs de modèle dans les sorties : Chaque variable dépendante indiquée dans l'onglet Variables génère un modèle estimé séparé. Les modèles sont différenciés par des noms uniques constitués d'un préfixe personnalisable et d'un suffixe entier. Vous pouvez entrer un préfixe ou conserver le préfixe par défaut du *modèle*.

Nombre maximum de décalages affichés dans les sorties ACF et PACF : Vous pouvez définir le nombre maximum de décalages affichés dans les tableaux et diagrammes d'autocorrélations et d'autocorrélations partielles.

Fonctions supplémentaires de la commande TSMODEL

Vous pouvez personnaliser la modélisation de votre série chronologique en collant vos sélections dans une fenêtre de syntaxe, puis en modifiant la syntaxe de la commande TSMODEL qui en résulte. Le langage de syntaxe de commande vous permet de :

- Indiquer la période saisonnière des données (avec le mot de passe SEASONLENGTH de la sous-commande AUXILIARY). Cette opération remplace la périodicité courante (le cas échéant) pour le jeu de données actif.
- Indiquer les décalages non consécutifs pour ARIMA personnalisé et les composants de la fonction de transfert (avec les sous-commandes ARIMA et TRANSFERFUNCTION). Par exemple, vous pouvez spécifier un modèle ARIMA personnalisé avec des décalages autorégressifs des ordres 1, 3 et 6 ou une fonction de transfert avec les décalages de numérateur des ordres 2, 5 et 8.
- Fournir plus d'un ensemble de spécifications de modélisation (par exemple, la méthode de modélisation, les ordres ARIMA, les variables indépendantes, etc.) pour une exécution unique de la procédure Time Series Modeler (avec la sous-commande MODEL).

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Appliquer des modèles de séries temporelles

La procédure Appliquer des modèles de séries chronologiques charge des modèles de séries chronologiques existants à partir d'un fichier externe et les applique à le jeu de données actif. Vous pouvez utiliser cette procédure pour obtenir des prévisions pour des séries pour lesquelles de nouvelles données ou des données révisées sont disponibles, sans avoir à recréer les modèles. Models are generated using the Time Series Modeler procedure.

Exemple : Vous êtes gestionnaire de stock pour un distributeur mondial et êtes responsable de 5 000 produits. Vous utilisez le modélisateur expert pour créer des modèles qui calculent les prévisions des ventes sur les trois mois à venir pour chaque produit. Votre entrepôt de données est actualisé tous les mois avec les données de vente réelles que vous souhaitez utiliser pour produire des prévisions actualisées tous les mois. La procédure Appliquer des modèles de séries chronologiques vous

permet d'accomplir ceci en utilisant des modèles originaux et de réestimer facilement les paramètres des modèles pour représenter les nouvelles données.

Statistiques : Mesures de la qualité de l'ajustement : R -carré stationnaire, R -carré (R^2), erreur quadratique moyenne (RMSE), erreur absolue moyenne (MAE), erreur de pourcentage absolue moyenne (MAPE), erreur absolue maximum (MaxAE), erreur de pourcentage absolue maximum (MaxAPE) et critère d'information bayésien normalisé (BIC). Résidus : fonction d'autocorrélation, fonction d'autocorrélation partielle, Q de Ljung-Box.

Tracés : Tracés récapitulatifs parmi tous les modèles : histogrammes de R -carré stationnaire, R -carré (R^2), erreur quadratique moyenne (RMSE), erreur absolue moyenne (MAE), erreur de pourcentage absolue moyenne (MAPE), erreur absolue maximum (MaxAE), erreur de pourcentage absolue maximum (MaxAPE), critère d'information bayésien normalisé (BIC) ; boîtes à moustaches d'autocorrélations résiduelles et partielles. Résultats des modèles individuels : valeurs de prévision, valeurs d'ajustement, valeurs observées, limites de confiance supérieure et inférieure, autocorrélations résiduelles et partielles.

Remarques sur les données Appliquer des modèles de séries chronologiques

Données : Les variables (indépendantes et dépendantes) auxquelles les modèles seront appliqués doivent être numériques.

Hypothèses : Les modèles sont appliqués à des variables dans le jeu de données actif ayant les mêmes noms que les variables indiquées dans le modèle. Toutes les variables de ce type sont traitées en tant que séries chronologiques, ce qui veut dire que chaque observation représente un point dans le temps, avec des observations successives séparées par un intervalle de temps constant.

- **Prévisions :** Pour produire des prévisions à l'aide de modèles avec des variables indépendantes (prédicteur), le jeu de données actif doit contenir des valeurs de ces variables pour toutes les observations de la période de prévision. Si les paramètres du modèle sont réestimés, les variables indépendantes ne doivent contenir aucune valeur manquante dans la période d'estimation.

Définition des dates

La procédure Appliquer des modèles de séries chronologiques requiert que la périodicité, le cas échéant, du jeu de données actif corresponde à la périodicité des modèles à appliquer. Si vous calculez les prévisions en utilisant simplement le même jeu de données (éventuellement avec des données nouvelles ou revues) que celui utilisé pour construire ce modèle, alors cette condition est remplie. S'il n'existe aucune périodicité pour le jeu de données actif, vous aurez la possibilité d'accéder à la boîte de dialogue Définir des dates pour en créer une. Si, cependant, les modèles ont été créés sans indiquer de périodicité, alors le jeu de données actif ne devrait pas en comporter non plus.

Pour appliquer des modèles

1. À partir des menus, sélectionnez :

Analyser > Prévision > Appliquer les modèles de causalité temporelle...

2. Enter the file specification for a model file or click **Browse** and select a model file (model files are created with the Time Series Modeler procedure).

En option, vous pouvez :

- Réestimer les paramètres de modèle en utilisant les données du jeu de données actif. Les prévisions sont créées en utilisant les paramètres réestimés.
- Enregistrer les prévisions, les intervalles de confiance et les résidus du bruit.
- Enregistrer les modèles réestimés au format XML.

Paramètres de modèle et mesures de la qualité de l'ajustement

Charger à partir du fichier de modèle : Les prévisions sont produites à l'aide des paramètres de modèle du fichier de modèle, sans réestimation de ces paramètres. Les mesures de la qualité de l'ajustement affichées dans la sortie et utilisées pour filtrer les modèles (les plus appropriés ou les moins appropriés) sont extraites du fichier de modèle et reflètent les données utilisées pour le développement de chaque modèle (ou pour leur dernière actualisation). Grâce à cette option, les prévisions ne prennent pas en compte les données historiques, pour les variables dépendantes comme pour les variables

indépendantes, dans le jeu de données actif. Vous devez choisir **Réestimer à partir des données** pour que les données historiques aient un impact sur les prévisions. En outre, les prévisions ne prennent pas en compte les valeurs des séries dépendantes dans la période de prévision, cependant elles prennent en compte celles des variables indépendantes. Si vous disposez de valeurs plus actuelles pour les séries dépendantes et que vous souhaitez les inclure dans les prévisions, vous devez réestimer la période d'estimation en l'ajustant pour y inclure ces valeurs.

Réestimer à partir des données : Les paramètres de modèle sont réestimés en utilisant les données du jeu de données actif. La réestimation des paramètres de modèle n'a aucun effet sur la structure de modèle. Par exemple, un modèle ARIMA(1,0,1) reste tel quel, mais les paramètres autorégressifs et de moyenne mobile seront réestimés. La réestimation ne résulte pas dans la détection de nouvelles valeurs extrêmes. Les valeurs extrêmes, s'il y en a, sont toujours extraites du fichier de modèle.

- **Période d'estimation** : La période d'estimation définit l'ensemble des observations utilisées pour réestimer les paramètres du modèle. Par défaut la période d'estimation inclut toutes les observations dans le jeu de données actif. Pour définir la période d'estimation, sélectionnez **Dans une plage de temps ou d'observations** dans la boîte de dialogue Sélectionner des observations. En fonction des données disponibles, la période d'estimation utilisée par la procédure peut varier selon le modèle et donc être différente de la valeur affichée. Pour un modèle donné, la vraie période d'estimation est la période restante après l'élimination de toutes les valeurs manquantes contiguës, à partir de la variable dépendante du modèle, qui se produit au début ou à la fin de la période d'estimation spécifiée.

Période de prévision

La période de prévision pour chaque modèle commence toujours par la première observation après la fin de la période d'estimation et va jusqu'à la dernière observation du jeu de données actif ou jusqu'à une date indiquée par l'utilisateur. Si les paramètres ne sont pas réestimés (ce qui se produit par défaut), alors la période d'estimation pour chaque modèle est l'ensemble des observations utilisées lors du développement du modèle (ou lors de sa dernière actualisation).

- **Première observation après la fin de la période d'estimation jusqu'à la dernière observation du jeu de données actif** : Sélectionnez cette option lorsque la fin de la période d'estimation est antérieure à la dernière observation du jeu de données actif et que vous souhaitez avoir des prévisions jusqu'à la dernière observation.
- **Première observation après la fin de la période d'estimation jusqu'à une date spécifique** : Sélectionnez cette option pour indiquer explicitement la fin de la période de prévision. Entrez les valeurs dans toutes les cellules de la grille Date.

Si aucune spécification de date n'a été définie pour le jeu de données actif, la grille Date affiche la colonne unique *Observation*. Pour indiquer la fin de la période de prévision, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

La colonne *Cycle* (si elle est présente) dans la grille Date se rapporte à la valeur de la variable *CYCLE_* du jeu de données actif.

Sortie

Les sorties disponibles incluent des résultats pour les modèles individuels ainsi que des résultats parmi tous les modèles. Les résultats pour les modèles individuels peuvent être limités à l'ensemble des modèles les plus ou les moins ajustés en fonction de critères définis par l'utilisateur.

Tableaux de statistiques et de prévision

L'onglet Statistiques fournit des options pour afficher des tableaux de statistiques d'ajustement du modèle, de paramètres de modèle, de fonctions d'autocorrélation et de prévisions. A moins que les paramètres de modèle ne soient réestimés (**Réestimer à partir des données** dans l'onglet Modèles), les valeurs affichées des mesures d'ajustement, les valeurs Ljung-Box et des paramètres de modèle sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation). Les informations de valeurs extrêmes sont toujours prises dans le fichier de modèle.

Afficher les mesures d'ajustement, la statistique Ljung-Box et le nombre de valeurs extrêmes

par modèle : Sélectionnez (cochez) cette option pour afficher un tableau contenant les mesures d'ajustement, la valeur Ljung-Box et le nombre de valeurs extrêmes sélectionnés pour chaque modèle.

Mesures d'ajustement : Vous pouvez sélectionner une ou plusieurs des valeurs suivantes pour les inclure dans le tableau contenant les mesures d'ajustement pour chaque modèle :

- *R*-carré stationnaire
- *R*-carré
- Erreur quadratique moyenne.
- Erreur absolue moyenne en pourcentage
- Erreur absolue moyenne.
- Erreur de pourcentage maximale
- Erreur absolue Maximum
- BIC normalisé

Statistiques de comparaison des modèles : Ce groupe d'options contrôle l'affichage des tableaux contenant des statistiques servant à tous les modèles. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Qualité de l'ajustement.** Tableau des statistiques récapitulatives et des percentiles pour *R*-square fixe, *R*-square, erreur quadratique moyenne quadratique, erreur absolue moyenne en pourcentage, erreur absolue moyenne, erreur absolue maximale en pourcentage, erreur absolue maximale et critère d'information bayésien normalisé.
- **Fonction d'autocorrélation (ACF) résiduelle :** Tableau des statistiques et percentiles récapitulatifs pour les autocorrélations résiduelles dans tous les modèles estimés. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).
- **Fonction d'autocorrélation partielle (PACF) résiduelle :** Tableau des statistiques et percentiles récapitulatifs pour les autocorrélations partielles résiduelles dans tous les modèles estimés. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).

Statistiques de modèles individuels : Ce groupe d'options contrôle l'affichage des tableaux contenant des informations détaillées pour chaque modèle. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Estimations des paramètres :** Affiche un tableau d'estimation des paramètres pour chaque modèle. Des tableaux distincts sont affichés pour les modèles de lissage exponentiel et ARIMA. Si des valeurs extrêmes existent, leurs estimations de paramètres sont également affichées dans un tableau distinct.
- **Fonction d'autocorrélation (ACF) résiduelle :** Affiche un tableau des autocorrélations résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).
- **Fonction d'autocorrélation partielle (PACF) résiduelle :** Affiche un tableau des autocorrélations partielles résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations partielles. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).

Afficher les prévisions : Affiche un tableau des prévisions du modèle et les intervalles de confiance pour chaque modèle.

Tracés

L'onglet Tracés fournit des options pour l'affichage de tracés de statistiques d'ajustement de modèle, de fonctions d'autocorrélation et de valeurs de séries, y compris les prévisions.

Tracés de comparaison des modèles

Ce groupe d'options contrôle l'affichage des tracés contenant des statistiques servant à tous les modèles. A moins que les paramètres de modèle ne soient réestimés (**Réestimer à partir des données** dans l'onglet Modèles), les valeurs affichées sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation). En outre, les tracés d'autocorrélation sont uniquement disponibles si les paramètres de modèle sont réestimés. Chaque option génère un tracé distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- R-carré stationnaire
- R-carré
- Erreur quadratique moyenne
- Erreur de pourcentage absolue moyenne
- Erreur Absolue Moyenne
- Erreur de pourcentage absolue maximum
- Erreur Absolue Maximum
- BIC normalisé
- Fonction d'autocorrélation (ACF) résiduelle
- Fonction d'autocorrélation partielle (PACF) résiduelle

Tracés de modèles individuels

Série. Sélectionnez (cochez) cette option pour obtenir des tracés des prévisions pour chaque modèle. Les valeurs observées, les valeurs d'ajustement, les intervalles de confiance des valeurs d'ajustement et les autocorrélations sont uniquement disponibles si les paramètres du modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles). Vous pouvez sélectionner un ou plusieurs des éléments suivants pour les insérer dans le tracé :

- **Valeurs observées** : Les valeurs observées des séries dépendantes.
- **Prévisions** : Les prévisions du modèle pour la période de prévision.
- **Valeurs d'ajustement** : Les prévisions du modèle pour la période d'estimation.
- **Intervalles de confiance pour les prévisions** : Les intervalles de confiance pour la période de prévision.
- **Intervalles de confiance pour les valeurs d'ajustement** : Les intervalles de confiance pour la période d'estimation.

Fonction d'autocorrélation (ACF) résiduelle : Affiche un tracé des autocorrélations résiduelles pour chaque modèle estimé.

Fonction d'autocorrélation partielle (PACF) résiduelle : Affiche un tracé des autocorrélations partielles résiduelles pour chaque modèle estimé.

Limitation des sorties aux modèles les plus ou les moins ajustés

L'onglet Filtre des sorties fournit des options pour restreindre à la fois des sorties de tableau ou de graphique à un sous-ensemble de modèles. Vous pouvez choisir de limiter des sorties aux modèles les plus ou les moins ajustés en fonction de critères d'ajustement que vous fournissez. Par défaut, tous les modèles sont inclus dans les sorties. A moins que les paramètres de modèle ne soient réestimés (**Réestimer à partir des données** dans l'onglet Modèles), les valeurs des mesures d'ajustement utilisées pour filtrer les modèles sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation).

Modèles les plus appropriés : Sélectionnez (cochez) cette option pour inclure les modèles les plus ajustés dans les sorties. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les moins ajustés. Dans ce cas, les sorties concernent aussi bien les modèles les moins ajustés que les modèles les plus ajustés.

- **Nombre fixe de modèles** : Indique que les résultats sont affichés pour les n modèles les plus ajustés. Si le nombre dépasse le nombre total de modèles, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles** : Indique que les résultats sont affichés pour les modèles avec des valeurs de qualité d'ajustement dans les n pourcentages supérieurs dans tous les modèles.

Modèles les moins ajustés : Sélectionnez (cochez) cette option pour inclure les modèles les moins ajustés dans les sorties. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les plus ajustés. Dans ce cas, les sorties sont constituées des modèles les plus ajustés ainsi que des modèles les moins ajustés.

- **Nombre fixe de modèles** : Indique que les résultats sont affichés pour les n modèles les moins ajustés. Si le nombre dépasse le nombre total de modèles, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles** : Indique que les résultats sont affichés pour les modèles avec des valeurs de qualité d'ajustement dans les n pourcentages inférieurs dans tous les modèles.

Mesure de la qualité de l'ajustement : Sélectionnez la mesure de la qualité d'ajustement à utiliser pour filtrer les modèles. La valeur par défaut est R -carré stationnaire.

Enregistrement des prévisions et des spécifications de modèle

L'onglet Enregistrer vous permet d'enregistrer des prévisions de modèle en tant que nouvelles variables dans le jeu de données actif et d'enregistrer les spécifications de modèle dans un fichier externe au format XML.

Enregistrer les variables : Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résidus en tant que nouvelles variables dans le jeu de données actif. Chaque modèle génère son propre ensemble de nouvelles variables. Les nouvelles observations sont ajoutées si la période de prévision s'étend au-delà de la longueur des séries de variables dépendantes associées au modèle. A moins que les paramètres de modèle soient réestimés (**Réestimer à partir des données** dans l'onglet Modèles), les prévisions et les limites de confiance sont uniquement créées pour la période de prévision. Choisissez d'enregistrer de nouvelles variables en cochant la case Enregistrer correspondant à chaque variable. Par défaut, aucune nouvelle variable n'est enregistrée.

- **Prévisions** : Les prévisions de modèle.
- **Limites de confiance inférieures** : Limites de confiance inférieures pour les prévisions.
- **Limites de confiance supérieures** : Limites de confiance supérieures pour les prévisions.
- **Résidus du bruit** : Les résidus de modèle. Lors de la transformation des variables dépendantes (par exemple Log népérien), ces derniers sont les résidus des séries transformées. Ce choix est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).
- **Préfixe de nom de variable** : Indiquez les préfixes à utiliser pour des nouveaux noms de variables ou conservez les préfixes par défaut. Les noms de variable sont constitués du préfixe, du nom de la variable dépendante correspondante et d'un identificateur de modèle. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables. Le préfixe doit se conformer aux règles pour les noms de variable valides.

Exporter un fichier de modèle : Les spécifications de modèle, contenant des paramètres réestimés et des statistiques d'ajustement sont exportés vers le fichier indiqué au format XML. Cette option est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).

- **Fichier XML** : Les spécifications de modèle sont enregistrées dans un fichier XML qui peut être utilisé avec les applications IBM SPSS.
- **Fichier PMML** : Les spécifications de modèle sont enregistrées dans un fichier XML compatible-PMML qui peut être utilisé avec les applications compatibles PMML, y compris les applications IBM SPSS.

Options

L'onglet Options vous permet de spécifier le traitement des valeurs manquantes, définir la largeur de l'intervalle de confiance et définir le nombre de décalage affiché pour les autocorrélations.

Valeurs manquantes de l'utilisateur : Ces options contrôlent le traitement des valeurs manquantes de l'utilisateur.

- **Traiter comme non valide :** Les valeurs manquantes de l'utilisateur sont traitées comme des valeurs système manquantes.
- **Traiter comme valide :** Les valeurs manquantes de l'utilisateur sont traitées comme des données valides.

Stratégie de valeur manquante : Les règles suivantes s'appliquent au traitement des valeurs manquantes (inclut les valeurs système manquantes et les valeurs manquantes de l'utilisateur traitées comme non valides) :

- Les observations avec des valeurs manquantes de variable dépendante qui se produisent au cours de la période d'estimation sont incluses dans le modèle. Le traitement spécifique de la valeur manquante dépend de la méthode d'estimation.
- Pour les modèles ARIMA, un avertissement est émis si le prédicteur a une valeur manquante pour la période d'estimation. Aucun modèle utilisant ce prédicteur ne sera réestimé.
- Si une variable indépendante a des valeurs manquantes pour la période d'estimation, la procédure émet un avertissement et effectue les prévisions dans la limite du possible.

Largeur de l'intervalle de confiance (%). Les intervalles de confiance sont calculés pour les prédictions du modèle et les autocorrélations résiduelles. Vous pouvez spécifier n'importe quelle valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Nombre maximum de décalages affichés dans les sorties ACF et PACF : Vous pouvez définir le nombre maximum de décalages affichés dans les tableaux et tracés d'autocorrélations et d'autocorrélations partielles. Cette option est uniquement disponible si les paramètres de modèle sont réestimés (**Réestimer à partir des données** dans l'onglet Modèles).

Fonctions supplémentaires de la commande TSAPPLY

Les fonctions supplémentaires sont disponibles si vous collez vos sélections dans une fenêtre de syntaxe et modifiez la syntaxe de la commande TSAPPLY. Le langage de syntaxe de commande vous permet de :

- Spécifier que seul un sous-ensemble des modèles dans un fichier de modèle doit s'appliquer à le jeu de données actif (avec les mots-clés DROP et KEEP dans la sous-commande MODEL).
- Appliquer des modèles de deux ou plusieurs fichiers de modèles à vos données (avec la sous-commande MODEL). Par exemple, un fichier de modèle peut contenir des modèles pour des séries représentant des ventes de référence et un autre des modèles pour des séries représentant les revenus.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Désaisonnalisation

La procédure Désaisonnalisation décompose une série en une composante saisonnière, une combinaison d'une composante de tendance et de cycle, et une composante d'erreur. Cette procédure représente l'application de la méthode rapport/moyenne mobile (Census Method I).

Exemple : Un scientifique souhaite analyser les mesures mensuelles du niveau d'ozone dans une station météorologique donnée. Son objectif est de savoir si les données affichent une tendance. Afin de découvrir une tendance effective, il doit d'abord représenter la variation des mesures due aux effets saisonniers. Ce scientifique peut utiliser la procédure Désaisonnalisation pour supprimer toute variation saisonnière systématique. La tendance peut ensuite être analysée sur une série désaisonnalisée.

Statistiques : Ensemble de facteurs saisonniers.

Remarques sur les données de la désaisonnalisation

Données : Les variables doivent être numériques.

Hypothèses : Les variables ne doivent contenir aucune donnée manquante intégrée. Au moins une composante de date périodique doit être définie.

Estimer des facteurs saisonniers

1. À partir des menus, sélectionnez :

Analyse > Prévisions > Désaisonnalisation...

2. Sélectionnez des variables dans la liste disponible, puis déplacez-les vers la liste des variables. Cette liste contient uniquement des variables numériques.

Type de modèle. La procédure Désaisonnalisation propose deux approches différentes pour modéliser les facteurs saisonniers : multiplicative ou additive.

- *Multiplicative.* Facteur par lequel les séries désaisonnalisées sont multipliées pour fournir les séries d'origine. Ainsi, les composantes saisonnières qui sont proportionnelles au niveau global des séries. Les observations dépourvues de variation saisonnière ont une composante de 1.
- *Additive.* Les ajustements saisonniers sont ajoutés aux séries désaisonnalisées pour obtenir les valeurs observées. Cet ajustement permet de filtrer les effets saisonniers d'une série dans le but de mettre en évidence d'autres caractéristiques pouvant être masquées par la composante saisonnière. Par exemple, les composantes saisonnières qui ne dépendent pas du niveau global de la série. Les observations sans variation saisonnière ont une composante saisonnière de 0.

Pondération de la moyenne mobile : Les options du groupe Pondération de la moyenne mobile vous permettent de spécifier le mode de traitement de la série lors du calcul des moyennes mobiles. Ces options sont uniquement disponibles si la périodicité de la série est régulière. Si la périodicité est impaire, tous les points sont pondérés de façon égale.

- *égal pour tous les points.* Les moyennes mobiles sont calculées avec un intervalle égal à la périodicité et avec tous les points pondérés de manière égale. Cette méthode est toujours utilisée si la périodicité est impaire.
- *points de fin pondérés par 0,5.* Les moyennes mobiles des séries dont la périodicité est paire sont calculées sur un intervalle égal à la périodicité plus un, les points de fin de l'intervalle étant pondérés par 0,5.

(En option) Vous avez également ces possibilités :

- Cliquer sur **Enregistrer** pour indiquer le mode d'enregistrement des nouvelles variables.

Enregistrement de la désaisonnalisation

Créer des variables : Vous permet de choisir le mode de traitement des nouvelles variables.

- *ajouter au fichier.* Les nouvelles séries créées par la procédure de désaisonnalisation sont enregistrées comme variables régulières dans votre jeu de données actif. Le nom des variables est formé d'un préfixe de trois lettres, d'un trait de soulignement et d'un nombre.
- *Remplacer les anciennes.* Les nouvelles séries créées par la procédure de désaisonnalisation sont enregistrées comme variables temporaires dans votre jeu de données actif. Simultanément, les variables temporaires existantes créées par les procédures Prévision sont ignorées. Le nom des variables est formé d'un préfixe à trois lettres, d'un signe dièse (#) et d'un nombre.
- *ne pas créer.* Les nouvelles séries ne sont pas ajoutées au de l'ensemble de données actif.

Noms des nouvelles variables

La procédure Désaisonnalisation crée quatre nouvelles variables (séries) avec les préfixes à trois lettres suivants, pour chaque série spécifiée :

SAF : *Facteurs de désaisonnalisation.* Ces valeurs indiquent l'effet de chaque période sur chaque niveau de la série.

SAS : *Séries désaisonnalisées*. Il s'agit des valeurs obtenues après la suppression de la variation saisonnière d'une série.

STC : *Composantes tendance-cycle lissées*. Ces valeurs indiquent la tendance et le comportement cyclique présents dans la série.

ERR : *Valeurs de résidu ou "d'erreur"*. Les valeurs restantes après que les composantes saisonnières, de tendance et de cycle ont été supprimées de la série.

Fonctions supplémentaires de la commande SEASON

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Indiquer une périodicité dans la commande SEASON au lieu de sélectionner l'une des solutions offertes par la procédure Définir des dates.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Tracés spectraux

La procédure Tracés spectraux permet d'identifier le comportement périodique des séries chronologiques. Au lieu d'analyser la variation d'un point à un autre, elle permet d'analyser la variation de la série dans son ensemble en tant que composantes périodiques des fréquences différentes. Les séries de lissage disposent de composantes périodiques plus fortes à basse fréquence ; la variation aléatoire (« white noise ») propage la force de ces composantes sur toutes les autres fréquences.

Les séries incluant des données manquantes ne peuvent pas être analysées avec cette procédure.

Exemple : La vitesse à laquelle les nouvelles maisons sont construites représente un baromètre important de l'état de l'économie. Les données relatives aux mises en chantier affichent généralement une composante saisonnière forte. Mais ces données contiennent-elles des cycles plus longs que les analystes doivent connaître lors de l'évaluation des chiffres en cours ?

Statistiques : Transformations sinus et cosinus, valeur du périodogramme et estimation de densité spectrale pour chaque composante de fréquence ou de période. Lorsque l'analyse bivariée est sélectionnée : parties réelles et imaginaires du périodogramme croisé, densité cospectrale, spectre en quadrature, gain, cohérence carrée et spectre de phase pour chaque composante de fréquence ou de période.

Tracés : Pour les analyses univariées et bivariées : périodogramme et densité spectrale. Pour les analyses bivariées : cohérence carrée, spectre en quadrature, amplitude croisée, densité cospectrale, spectre de phase et de gain.

Remarques sur les données du tracés spectraux

Données : Les variables doivent être numériques.

Hypothèses : Les variables ne doivent contenir aucune donnée manquante intégrée. La série chronologique à analyser doit être stationnaire et toute moyenne différente de zéro doit en être soustraite.

- *Stationnaire*. Condition qui doit être remplie par les séries temporelles auxquelles vous ajustez un modèle ARIMA. Les séries MA pures sont stationnaires, mais les séries AR et ARMA ne le sont pas toujours. Une série stationnaire a une moyenne et une variance constantes dans le temps.

Obtention d'une analyse spectrale

1. A partir des menus, sélectionnez :

Analyse > Séries temporelles > Analyse spectrale...

2. Sélectionnez des variables dans la liste disponible, puis déplacez-les vers la liste des variables. Cette liste contient uniquement des variables numériques.

3. Sélectionnez l'une des options Fenêtre spectrale pour choisir le mode de lissage du périodogramme, en vue d'obtenir une estimation de densité spectrale. Les options de lissage disponibles sont les suivantes : Tukey-Hamming, Tukey, Parzen, Bartlett, Daniell (unité) et Aucune.
- *Tukey-Hamming*. Les poids sont $W_k = .54D_p(2 \pi f_k) + .23D_p(2 \pi f_k + \pi/p) + .23D_p(2 \pi f_k - \pi/p)$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'étendue et F_p est le noyau de Dirichlet d'ordre p .
 - *Tukey*. Les poids sont $W_k = 0.5D_p(2 \pi f_k) + 0.25D_p(2 \pi f_k + \pi/p) + 0.25D_p(2 \pi f_k - \pi/p)$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'étendue et F_p est le noyau de Dirichlet d'ordre p .
 - *Parzen*. Les poids sont $W_k = 1/p(2 + \cos(2 \pi f_k)) (F_{[p/2]}(2 \pi f_k))^2$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'étendue et $F_{[p/2]}$ est le noyau de Fejer d'ordre $p/2$.
 - *Bartlett*. La forme d'une fenêtre spectrale pour laquelle les poids de la moitié supérieure de la fenêtre sont calculés comme suit : $W_k = F_p(2 \pi f_k)$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'étendue et F_p est le noyau de Fejer d'ordre p . La moitié inférieure est symétrique avec la moitié supérieure.
 - *Daniell (Unité)*. Forme d'une fenêtre spectrale pour laquelle les pondérations sont égales à 1.
 - *Néant*. Pas de lissage. Si cette option est sélectionnée, la densité spectrale estimée est identique à celle évaluée par le périodogramme.

Span. Plage de valeurs consécutives sur laquelle le lissage est effectué. Généralement, un entier impair est utilisé. Les grands intervalles appliquent à la densité spectrale du tracé un lissage plus important que les intervalles réduits.

centrer les variables. Ajuste la série de manière à obtenir une moyenne égale à zéro avant de calculer le spectre et d'éliminer le terme élevé qui peut être associé avec la moyenne de la série.

Analyse bivariée - chaque variable avec la première : Si vous avez sélectionné au moins deux variables, vous pouvez sélectionner cette option pour demander des analyses spectrales bivariées.

- La première variable de la liste Variable(s) est traitée comme une variable explicative et toutes les autres variables comme des variables dépendantes.
- Chaque série suivant la première est analysée avec la première série indépendamment de l'autre série nommée. Des analyses univariées de chaque série sont également lancées.

Tracé : Le périodogramme et la densité spectrale sont disponibles pour les analyses univariées et bivariées. Toutes les autres options sont disponibles uniquement pour les analyses bivariées.

- *périodogramme*. Tracé non lissé de l'amplitude spectrale (tracé sur une échelle logarithmique) par rapport à la fréquence ou à la période. De faibles variations caractérisent une série lissée. Des variations étendues sur tous les effectifs indiquent un "bruit blanc".
- *cohérence carrée*. Produit des gains des deux séries.
- *Spectre en quadrature*. Partie imaginaire d'un périodogramme croisé qui correspond à la mesure de la corrélation entre les composantes de fréquence en phase de deux séries temporelles. Les composantes sont décalées de $\pi/2$ radians.
- *amplitude croisée*. Racine carrée de la somme du carré de la densité cospectrale et du carré du spectre en quadrature.
- *Densité spectrale*. Périodogramme qui a été lissé pour en éliminer les variations irrégulières.
- *densité cospectrale*. Partie réelle d'un périodogramme croisé qui correspond à la mesure de la corrélation entre les composants des effectifs en phase de deux séries temporelles.
- *spectre de phase*. Mesure de l'étendue pour laquelle chaque composante d'effectif d'une série succède ou précède une autre.
- *gain*. Quotient de la division de l'amplitude croisée par la densité spectrale de l'une des séries. Chacune des deux séries dispose de sa propre valeur de gains.

par fréquence. Tous les tracés sont créés par fréquence, de la fréquence 0 (constante ou moyenne) à la fréquence 0,5 (terme correspondant à un cycle de deux observations).

par période. Tous les tracés sont produits par période, allant de 2 (cycle de deux observations) à une période égale au nombre d'observations (constante ou moyenne). La période est affichée sur une échelle logarithmique.

Fonctions supplémentaires de la commande SPECTRA

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Enregistrer les variables d'analyse spectrale calculées dans le jeu de données actif en vue d'une prochaine utilisation.
- Spécifier les pondérations personnalisées pour la fenêtre spectrale.
- Produire les tracés par fréquence et par période.
- Imprimer un listage complet de chaque valeur affichée dans le tracé.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Modèles de causalité temporelle

La modélisation de causalité temporelle tente de découvrir des relations de causalité clés dans les données de séries temporelles. Dans la modélisation de causalité temporelle, vous spécifiez un ensemble de séries cibles et un ensemble d'entrées candidates sur ces cibles. La procédure crée ensuite un modèle de série temporelle autorégressif pour chaque cible et inclut uniquement les entrées ayant une relation de causalité avec la cible. Cette approche diffère de la modélisation traditionnelle de série temporelle où vous devez spécifier explicitement les prédicteurs pour la série cible. Vu que la modélisation de causalité temporelle implique généralement la construction de modèles pour plusieurs séries temporelles associées, le résultat est dénommé *système modèle*.

Dans le contexte de modélisation de causalité temporelle, le terme *causal* se réfère à la causalité au sens de Granger. Une série temporelle X cause au sens de Granger une autre série temporelle Y si la régression pour Y sur les valeurs passées de X et de Y génère un meilleur modèle pour Y que la régression uniquement sur les valeurs passées de Y.

Exemples

Les décideurs métier peuvent utiliser la modélisation de causalité temporelle pour identifier des relations de causalité dans un vaste ensemble de métriques basées temps qui décrivent l'activité. L'analyse peut révéler un petit nombre d'entrées contrôlables qui ont l'impact le plus déterminant sur les indicateurs de performances clés.

Les responsables de systèmes informatiques à grande échelle peuvent utiliser la modélisation de causalité temporelle pour détecter des anomalies dans un large éventail de métriques opérationnelles interdépendantes. Le modèle de causalité permet ensuite d'aller au-delà de la détection d'anomalies et de découvrir les causes premières les plus probables des anomalies.

Exigences relatives aux champs

Au moins une cible doit être spécifiée. Par défaut, les champs ayant le rôle prédéfini Aucun ne sont pas utilisés.

Structure de données

La modélisation de causalité temporelle prend en charge deux types de structures de données.

Données par colonnes

Pour les données basées colonne, chaque zone de série temporelle contient les données d'une seule série temporelle. Il s'agit de la structure traditionnelle des données de série temporelle, telle qu'utilisée par Time Series Modeler.

Données multidimensionnelles

Pour les données multidimensionnelles, chaque champ de série temporelle contient les données de plusieurs séries temporelles. Des séries temporelles distinctes au sein d'un champ donné sont ensuite identifiées par un ensemble de valeurs de champs catégoriels appelés champs de *dimension*. Par exemple, les données de ventes de deux circuits de vente différents (détail et Web) pourraient être stockées dans une zone *ventes* unique. Une zone de dimension nommée *canal*, avec pour valeurs 'détail' et 'Web', identifie les enregistrements associés à chacun des deux circuits de vente.

Obtention d'un modèle de causalité temporelle

Cette fonction nécessite l'option Statistics Forecasting.

À partir des menus, sélectionnez :

Analyser > Prévision > Créer des modèles de causalité temporelle...

1. Si les observations sont définies par un champ de date/heure, renseignez ce champ.
2. Avec des données multidimensionnelles, spécifiez les champs de dimension qui identifient la série temporelle.
 - L'ordre des champs de dimension spécifiés définit l'ordre dans lequel ils apparaissent dans toutes les boîtes de dialogue et sorties suivantes. Utilisez les flèches de défilement vers le haut et vers le bas pour redéfinir l'ordre des champs de dimension.
 - Vous pouvez spécifier des sous-ensembles de valeurs de dimension afin de limiter l'analyse à des valeurs spécifiques des champs de dimension. Par exemple, si vous avez des dimensions de région (region) et de marque (brand), vous pouvez limiter l'analyse à une région spécifique. Les sous-ensembles de dimension s'appliquent à tous les champs de métrique qui sont utilisés dans l'analyse.
 - Vous pouvez également personnaliser l'analyse en spécifiant des valeurs de dimension par rôle au niveau des champs de métrique. Par exemple, si vous avez une dimension Sales Channel (avec les valeurs 'retail' et 'web') et des métriques pour Sales et Advertising dans ces dimensions, vous pouvez spécifier "web advertising" comme entrée pour les ventes de type retail et web. Par défaut, ce type de personnalisation est activé et limité à la sélection d'un nombre spécifié de valeurs distinctes (par défaut, 250) de chaque champ de dimension.
3. Cliquez sur **Continuer**

Remarque : Les étapes 1 2 et 3 ne s'appliquent pas si le jeu de données actif comporte une spécification de date. Les spécifications de date sont créées à partir de la boîte de dialogue de définition de dates ou de la commande DATE.
4. Cliquez sur **Champs** pour indiquer la série temporelle à inclure dans le modèle et spécifier comment les observations sont définies. Au moins un champ doit être spécifié comme cible ou comme entrée et cible.
5. Cliquez sur **Spécifications de données** pour spécifier des paramètres facultatifs qui incluent l'intervalle de temps pour les fonctions d'analyse, d'agrégation et de distribution et le traitement des valeurs manquantes.
6. Cliquez sur **Options de génération** pour définir la période d'estimation, indiquer le contenu de la sortie et spécifier des paramètres de génération tels que le nombre maximal d'entrées par cible.
7. Cliquez sur **Options de modèle** pour demander des prévisions, les enregistrer et exporter le système modèle dans un fichier externe.
8. Cliquez sur **Exécuter** pour exécuter la procédure.

Série temporelle à modéliser

Sur l'onglet Champs, utilisez les paramètres **Série temporelle** pour spécifier la série à inclure dans le système de modèle.

Pour les données basées sur les colonnes, le terme *série* a la même signification que le terme *champ*. Pour les données multidimensionnelles, les champs contenant des séries temporelles sont appelés

champs de *métrique*. Pour les données multidimensionnelles, une série temporelle est définie par un champ de métrique et une valeur pour chacun des champs de dimension. Les considérations suivantes s'appliquent à la fois aux données basées sur les colonnes et aux données multidimensionnelles :

- Les séries pour lesquelles la valeur Entrées candidates ou A la fois cible et entrée est spécifiée sont incluses dans le modèle de chaque cible. Le modèle de chaque cible inclut toujours des valeurs décalées de la cible elle-même.
- Les séries pour lesquelles la valeur Entrées forcées est spécifiée sont toujours incluses dans le modèle de chaque cible.
- Au moins une série doit être spécifiée comme cible ou comme cible et entrée.
- Lorsque l'option **Utiliser des rôles prédéfinis** est sélectionnée, les champs dont le rôle est Entrée sont définis en tant qu'entrées candidates. Aucun rôle prédéfini n'est mappé à une entrée forcée.

Données multidimensionnelles

Pour les données multidimensionnelles, vous spécifiez des champs de métrique et les rôles qui leur sont associés dans la grille où chaque ligne spécifie une métrique et un rôle uniques. Par défaut, le système de modèle inclut des séries pour toutes les combinaisons des champs de dimension de chaque ligne de la grille. Supposons qu'il existe des dimensions *region* et *brand*, par défaut, si la métrique *sales* est spécifiée en tant que cible, cela signifie qu'il existe une série cible *sales* distincte pour chaque combinaison d'observations *region* et *brand*.

Pour chaque ligne de la grille, vous pouvez personnaliser l'ensemble de valeurs de n'importe quel champ de dimension en cliquant sur les points de suspension correspondant à une dimension. Cette action ouvre la sous-boîte de dialogue Sélectionner des valeurs de dimension. Vous pouvez également ajouter, supprimer ou copier des lignes de grille.

La colonne **Comptage des séries** affiche le nombre d'ensembles de valeurs de dimension actuellement spécifiés pour la métrique associée. La valeur affichée peut être supérieure au nombre réel de séries (une série par ensemble). Cette condition se produit lorsque certaines des combinaisons de valeurs de dimension spécifiées ne correspondent pas aux séries contenues par la métrique associée.

Sélectionnez des valeurs pour Dimension

Pour les données multidimensionnelles, vous pouvez personnaliser l'analyse en spécifiant les valeurs de dimension qui s'appliquent à un champ de métrique spécifique avec un rôle donné. Par exemple, si *ventes* est une zone de métrique et que *canal* est une dimension pouvant recevoir comme valeur 'vente au détail' et 'Web', vous pouvez spécifier que les ventes 'Web' sont une entrée et les ventes 'au détail' une cible. Vous pouvez également spécifier des sous-ensembles de dimension qui s'appliquent à tous les champs de métrique utilisés dans l'analyse. Par exemple, si *region* est un champ de dimension qui indique une région géographique, vous pouvez limiter l'analyse à des régions particulières.

Toutes les valeurs

Indique que toutes les valeurs du champ de dimension en cours sont incluses. Il s'agit de l'option par défaut.

Sélectionnez les valeurs à inclure ou à exclure

Cette option vous permet de spécifier l'ensemble de valeurs du champ de dimension en cours.

Lorsque **Inclure** est sélectionné pour le **Mode**, seules les valeurs spécifiées dans la liste **Valeurs sélectionnées** sont incluses. Quand **Exclure** est sélectionné pour **Mode**, toutes les valeurs autres que celles spécifiées dans la liste **Valeurs sélectionnées** sont incluses.

Vous pouvez filtrer l'ensemble de valeurs dans lequel effectuer une sélection. Les valeurs conformes à la condition de filtrage apparaissent dans l'onglet **Correspondance** et celles qui n'y correspondent pas figurent dans l'onglet **Sans correspondance** de la liste **Valeurs non sélectionnées**. L'onglet **Toutes** recense toutes les valeurs non sélectionnées, sans considération de condition de filtrage quelconque.

- Vous pouvez utiliser des astérisques (*) pour indiquer des caractères génériques lorsque vous spécifiez un filtre.

- Pour désélectionner le filtre en cours, spécifiez une valeur vide pour le terme recherché sur la boîte de dialogue Filtrer les valeurs affichées.

Observations

Sur l'onglet Champs, utilisez les paramètres **Observations** pour spécifier les champs qui définissent les observations.

Remarque : Si le jeu de données actif comporte une spécification de date, les observations sont définies par la spécification de date et ne peuvent pas être modifiées dans la procédure de modélisation de causalité temporelle. Les spécifications de date sont créées à partir de la boîte de dialogue de définition de dates ou de la commande DATE.

Observations définies par date/heures

Vous pouvez spécifier des observations définies par un champ avec un format de date, d'heure, ou de date/heure, ou par une champ de chaîne qui représente une date/heure. Les zones de chaîne peuvent représenter un date au format AAAA-MM-JJ, une heure au format HH:MM:SS, ou une date/heure au format AAAA-MM-JJ HH:MM:SS. Les zéros non significatifs peuvent être omis dans la représentation sous forme de chaîne. Par exemple, la chaîne 2014-9-01 équivaut à 2014-09-01.

Outre le champ qui définit les observations, sélectionnez l'intervalle de temps approprié qui décrit les observations. En fonction de l'intervalle de temps donné, vous pouvez également spécifier d'autres paramètres, tels que l'intervalle compris entre des observations (incrémentations) ou le nombre de jours par semaine. Les considérations suivantes s'appliquent à l'intervalle de temps :

- Utilisez la valeur **Irrégulier** lorsque les observations ne sont pas régulièrement espacées dans le temps, telles que l'heure à laquelle un bon de commande est traité. Lorsque la valeur **Irrégulier** est sélectionnée, vous devez spécifier l'intervalle de temps utilisé pour l'analyse à partir des paramètres **Intervalle de temps** de l'onglet Spécifications de données.
- Lorsque les observations représentent une date et une heure et que l'intervalle de temps est exprimé en heures, en minutes ou en secondes, utilisez les paramètres **Nombre d'heures par jours**, **Nombre de minutes par jour** ou **Nombre de secondes par jour**. Lorsque les observations représentent une heure (durée) qui ne fait référence à aucune date et que l'intervalle de temps est exprimé en heures, en minutes ou en secondes, utilisez les paramètres **Heures (non périodiques)**, **Minutes (non périodiques)** ou **Secondes (non périodiques)**.
- Selon l'intervalle de temps sélectionné, la procédure peut détecter des observations manquantes. La détection d'observations manquantes est nécessaire vu que la procédure suppose que toutes les observations sont également espacées dans le temps et qu'il n'y a pas d'observations manquantes. Par exemple, si l'intervalle de temps est exprimé en jours et que la date 2014-10-27 est suivie de la date 2014-10-29, il existe une observation manquante pour la date 2014-10-28. Des valeurs sont imputées à toutes les observations manquantes. Les paramètres de gestion des valeurs manquantes peuvent être spécifiés dans l'onglet Spécifications des données.
- L'intervalle de temps spécifié permet à la procédure de détecter des observations multiples au sein du même intervalle qui doivent être agrégées et d'aligner les observations par rapport à la limite de l'intervalle (par exemple, le premier du mois) pour garantir que les observations soient également espacées. Par exemple, si l'intervalle de temps est exprimé en mois (Months), plusieurs dates d'un même mois sont agrégées. Ce type d'agrégation est appelé *regroupement*. Par défaut, les observations sont additionnées lorsqu'elles sont regroupées. Vous pouvez spécifier une autre méthode de regroupement, telle que la moyenne des observations, à partir des paramètres **Agrégation et distribution** sur l'onglet Spécifications de données.
- Pour certains intervalles de temps, les paramètres supplémentaires peuvent définir une coupure dans les intervalles normalement avec espacement uniforme des observations. Par exemple, si l'intervalle de temps est toujours exprimé en jours, mais que seuls les jours de la semaine sont valides, vous pouvez indiquer qu'une semaine comporte cinq jours et qu'elle commence le lundi.

Observations définies par périodes ou par périodes cycliques

Des observations peuvent être définies par un ou plusieurs champs de type entier qui représentent des périodes ou des cycles de répétition de périodes, jusqu'à un nombre arbitraire de niveaux de

cycles. Via cette structure, vous pouvez décrire des séries d'observations qui ne se prêtent pas à un intervalle de temps standard. Par exemple, un exercice financier de 10 mois seulement peut être décrit à l'aide d'un champ de cycle qui représente les exercices et un champ de période qui représente les mois, la longueur d'un cycle étant égale à 10.

Les champs qui spécifient des périodes cycliques définissent une hiérarchie de niveaux périodiques où le niveau le plus bas est défini par le champ **Période**. Le niveau juste au-dessus est spécifié par un champ cycle dont le niveau est 1, suivi par un champ cycle dont le niveau est 2, et ainsi de suite. Les valeurs de zone de chaque niveau, sauf la plus élevée, doivent être périodiques par rapport au niveau supérieur suivant. Les valeurs du niveau le plus haut ne peuvent pas être périodiques. Par exemple, dans le cas d'un exercice financier de 10 mois, les mois sont périodiques au sein des exercices et ceux-ci ne sont pas périodiques.

- La longueur d'un cycle à un niveau spécifique est la périodicité du niveau suivant le plus bas. Dans le cas de l'exercice financier, il existe un seul niveau de cycle et la longueur du cycle est égale à 10 puisque le niveau suivant le plus bas représente les mois et l'exercice financier spécifié comporte 10 mois.
- Spécifiez la valeur de départ pour tout champ périodique qui ne démarre pas à partir de 1. Ce paramètre est nécessaire pour détecter les valeurs manquantes. Par exemple, si un champ périodique débute à partir de 2, mais que la valeur de début spécifiée est 1, la procédure considère qu'il existe une valeur manquante pour la première période de chaque cycle de ce champ.

Observations définies par l'ordre des enregistrements

Pour les données basées sur les colonnes, vous pouvez spécifier que les observations sont définies par un ordre d'enregistrement où le premier enregistrement représente la première observation, le deuxième enregistrement représente la deuxième observation, et ainsi de suite. On considère alors que les enregistrements représentent des observations régulièrement espacées dans le temps.

Intervalle de temps pour l'analyse

L'intervalle de temps utilisé pour l'analyse peut être différent de celui des observations. Par exemple, si l'intervalle de temps des observations est exprimé en jours, vous pouvez choisir d'exprimer l'intervalle de temps des analyses en mois. Les données sont alors agrégées de données quotidiennes en données mensuelles avant la construction du modèle. Vous pouvez également choisir de distribuer les données d'un intervalle de temps plus long sur un intervalle de temps plus court. Par exemple, si les observations sont trimestrielles, vous pouvez les distribuer sur une base mensuelle.

Les options disponibles afin de définir l'intervalle de temps pour l'analyse dépendent de la manière dont les observations sont définies et de l'intervalle de temps de ces observations. Plus spécifiquement, lorsque les observations sont définies par des périodes cycliques ou lorsqu'une spécification de date est définie pour le jeu de données actif, seule l'agrégation est prise en charge. Dans ce cas, l'intervalle d'analyse doit être supérieur ou égal à celui des observations.

L'intervalle de temps pour l'analyse est spécifié dans les paramètres **Intervalle de temps** sur l'onglet Spécifications des données. La méthode utilisée pour agréger ou distribuer les données est spécifiée dans les paramètres **Agrégation et distribution** sur l'onglet Spécifications de données.

Agrégation et distribution

Fonctions d'agrégation

Lorsque l'intervalle de temps qui est utilisé pour l'analyse est plus long que celui des observations, les données d'entrée sont agrégées. Par exemple, une agrégation est effectuée lorsque l'intervalle de temps des observations est exprimé en jours et que l'intervalle de temps de l'analyse est exprimé en mois. Les fonctions d'agrégation suivantes sont disponibles : Moyenne, Somme, Mode, Min ou Max.

Fonctions de distribution

Lorsque l'intervalle de temps qui est utilisé pour l'analyse est plus court que celui des observations, les données d'entrée sont distribuées. Par exemple, une distribution est effectuée lorsque l'intervalle de temps des observations est exprimé en trimestres et que l'intervalle de temps de l'analyse est exprimé en mois. Les fonctions de distribution suivantes sont disponibles : Moyenne ou Somme.

Fonctions de regroupement

Le regroupement s'applique lorsque des observations sont définies par des dates et des heures et que plusieurs observations se produisent au cours du même intervalle de temps. Par exemple, si l'intervalle de temps des observations est exprimé en mois, plusieurs dates d'un même mois sont regroupées et associées au mois au cours duquel ces observations se produisent. Les fonctions de regroupement suivantes sont disponibles : Moyenne, Somme, Mode, Min ou Max. Le regroupement est toujours effectué lorsque les observations sont définies par des dates et des heures et que leur intervalle de temps est spécifié comme étant irrégulier.

Remarque : Bien que le regroupement soit une forme d'agrégation, il est effectué avant le traitement des valeurs manquantes. En revanche, l'agrégation formelle est effectuée après le traitement des valeurs manquantes. Lorsque l'intervalle de temps des observations est spécifié comme étant irrégulier, l'agrégation est effectuée uniquement à l'aide de la fonction de regroupement.

Agréger les observations de la journée à la journée précédente

Indique si des observations dont les heures chevauchent deux journées sont agrégées avec les valeurs de la journée précédente. Par exemple, pour des observations horaires avec une journée de huit heures qui commence à 20 heures, ce paramètre spécifie si les observations qui se produisent entre 00:00 et 04:00 sont incluses dans les résultats agrégés relatifs au jour précédent. Ce paramètre s'applique uniquement si l'intervalle de temps des observations est exprimé en heures par jour, en minutes par jour ou en secondes par jour et que l'intervalle de temps de l'analyse est exprimé en jours.

Paramètres personnalisés pour les champs spécifiés

Vous pouvez spécifier les fonctions d'agrégation, de distribution et de regroupement champ par champ. Ces paramètres remplacent les paramètres définis par défaut pour les fonctions d'agrégation, de distribution et de regroupement.

Valeurs manquantes

Les valeurs manquantes dans les données d'entrée sont remplacées par une valeur imputée. Les méthodes de remplacement suivantes sont disponibles :

Interpolation linéaire

Remplace les valeurs manquantes au moyen d'une interpolation linéaire. La dernière valeur valide avant la valeur manquante et la première valeur valide après la valeur manquante sont utilisées pour l'interpolation. Si la première ou la dernière observation de la série a une valeur manquante, les deux valeurs non manquantes les plus proches au début ou à la fin de la série sont utilisées.

Moyenne série

Remplace les valeurs manquantes par la moyenne de toute la série.

Moyenne des points voisins

Remplace les valeurs manquantes par la moyenne des valeurs valides qui les entourent. La sphère des points voisins est le nombre de valeurs valides au-dessus et au-dessous de la valeur manquante qui sont utilisées pour calculer la moyenne.

Valeur médiane des points voisins

Remplace les valeurs manquantes par la médiane des valeurs valides qui les entourent. La sphère des points voisins est le nombre de valeurs valides au-dessus et au-dessous de la valeur manquante qui sont utilisées pour calculer la valeur médiane.

Tendance linéaire

Cette option utilise toutes les observations non manquantes dans la série afin d'alimenter un modèle de régression linéaire simple, qui est ensuite utilisé pour imputer les valeurs manquantes.

Autres paramètres :

Pourcentage maximal de valeurs manquantes (%)

Spécifie le pourcentage maximal de valeurs manquantes autorisé pour une série. Les séries avec un nombre de valeurs manquantes supérieur à ce maximum sont exclues de l'analyse.

Les valeurs manquantes de l'utilisateur

Cette option spécifie si les valeurs manquantes de l'utilisateur sont traitées comme des données valides et par conséquent, incluses dans les séries. Par défaut, les valeurs manquantes de l'utilisateur sont exclues et traitées comme des valeurs système manquantes, lesquelles sont ensuite attribuées.

Options générales pour les données

Nombre maximal de valeurs distinctes par zone de dimension

Ce paramètre s'applique aux données multidimensionnelles et spécifie le nombre maximal de valeurs distinctes permises pour un champ de dimension quelconque. Par défaut, cette limite est fixée à 1 0000, mais elle peut être augmentée et atteindre un nombre arbitrairement élevé.

Options générales de construction

Largeur de l'intervalle de confiance (%)

Ce paramètre contrôle les intervalles de confiance pour les prévisions et pour les paramètres du modèle. Vous pouvez spécifier n'importe quelle valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Nombre maximal d'entrées pour chaque cible

Ce paramètre spécifie le nombre maximal d'entrées autorisé dans le modèle pour chaque cible. Vous pouvez spécifier un entier sur la plage 1 à 20. Le modèle de chaque cible inclut toujours des valeurs décalées d'elle-même, par conséquent, si la valeur 1 est affectée à ce paramètre, la seule sortie est la cible elle-même.

Tolérance du modèle

Ce paramètre contrôle le processus itératif qui est utilisé pour déterminer l'ensemble d'entrées le plus adapté pour chaque cible. Vous pouvez spécifier n'importe quelle valeur supérieure à zéro. La valeur par défaut est 0.001.

Seuil de valeur extrême (%)

Une observation est marquée comme étant une valeur extrême si la probabilité, calculée depuis le modèle, qu'il s'agisse d'une valeur extrême dépasse ce seuil. Vous pouvez spécifier une valeur comprise entre 50 et 100.

Nombre de décalages pour chaque entrée

Ce paramètre spécifie le nombre de termes de décalage pour chaque entrée pour chaque cible dans le modèle. Par défaut, le nombre de termes décalés est déterminé automatiquement à partir de l'intervalle de temps qui est utilisé pour l'analyse. Par exemple, si l'intervalle de temps est défini à Mois (avec un incrément d'un mois), le nombre de décalages est alors de 12. Vous pouvez aussi spécifier explicitement le nombre de décalages. La valeur spécifiée doit être un nombre entier compris entre 1 et 20.

Séries à afficher

Ces options spécifient les séries (cibles ou entrées) pour lesquelles une sortie est affichée. Le contenu de la sortie pour les séries spécifiées est déterminé par les paramètres **Options de sortie**.

Afficher les cibles associées aux modèles les plus adaptés

Par défaut, la sortie est affichée pour les cibles qui sont associées aux 10 modèles les plus adaptés, comme l'indique la valeur R-deux. Vous pouvez spécifier un nombre fixe différent ou un pourcentage pour les modèles les plus adaptés. Vous pouvez également sélectionner l'une des mesures suivantes d'adéquation des modèles :

R-deux

Mesure de qualité d'ajustement d'un modèle linéaire, parfois appelée coefficient de détermination. Il s'agit de la proportion de la variation de la variable cible, expliquée par le modèle. Cela va de 0 à 1. Les petites valeurs indiquent que le modèle ne correspond pas bien aux données.

Pourcentage d'erreur quadratique moyenne

Mesure illustrant l'ampleur de l'écart entre les valeurs prévues par le modèle et les valeurs observées dans les séries. Elle est indépendante des unités utilisées et peut donc servir pour comparer les séries comportant des unités différentes.

Erreur quadratique moyenne.

La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la série dépendante.

code identificateur de banque

Critère d'information bayésien. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. Le critère BIC (Bayesian Information Criterion) "pénalise" lui aussi les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais de façon plus stricte que le critère AIC (Akaike's Information Criterion).

AIC

Critère d'information d'Akaike. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. L'AIC "pénalise" les modèles surparamétrés (modèles complexes avec un grand nombre d'entrées, par exemple).

Spécifier des séries individuelles

Vous pouvez spécifier des séries individuelles pour lesquelles générer une sortie.

- Pour les données basées colonnes, spécifiez les champs contenant les séries voulues. L'ordre des champs spécifiés détermine leur ordre d'apparition dans la sortie.
- Pour les données multidimensionnelles; spécifiez une série spécifique en ajoutant une entrée dans la grille pour le champ de métrique qui contient la série. Indiquez ensuite les valeurs des champs de dimension qui définissent la série.
 - Vous pouvez entrer la valeur de chaque champ de dimension directement dans la grille ou vous pouvez effectuer une sélection dans la liste de valeurs de dimension disponibles. Pour effectuer une sélection dans cette liste, cliquez sur les points de suspension dans la cellule correspondant à la dimension voulue. Cette action ouvre la sous-boîte de dialogue Sélectionner des valeurs de dimension.
 - Vous pouvez effectuer une recherche dans la liste des valeurs de dimension depuis la sous-boîte de dialogue Sélectionnez une valeur pour Dimension en cliquant sur l'icône en forme de jumelles et en spécifiant un terme de recherche. Les espaces sont considérés faire partie du terme de recherche. Les astérisques (*) utilisés dans le terme recherché ne représentent pas des caractères génériques.
 - L'ordre des séries dans la grille détermine leur ordre d'apparition dans la sortie.

Pour les données basées sur les colonnes et les données multidimensionnelles, la sortie est limitée à 30 séries. Cette limite inclut les séries individuelles (entrées ou cibles) que vous spécifiez, ainsi que les cibles associées aux modèles les plus adaptés. Les séries spécifiées individuellement sont prioritaires sur les cibles associées aux modèles les plus adaptés.

Options de sortie

Ces options spécifient le contenu de la sortie. Les options du groupe **Sortie pour cibles** génèrent une sortie pour les cibles associées avec les modèles les plus adaptés dans les paramètres **Séries à afficher**. Les options du groupe **Sortie pour séries** génèrent une sortie pour les séries individuelles spécifiées dans les paramètres **Séries à afficher**.

Système de modèle global

Affiche une représentation graphique des relations de causalité entre les séries dans le système modèle. Les tableaux contenant les statistiques et les valeurs extrêmes de la qualité de l'ajustement pour les cibles affichées sont inclus dans le cadre de l'élément de sortie. Lorsque cette option est

sélectionnée dans le groupe **Sortie pour les séries**, un élément de sortie distinct est créé pour chaque série individuelle qui est spécifiée dans les paramètres **Séries à afficher**.

Les relations de causalité ont un niveau d'importance associé, où un niveau plus faible indique une connexion plus significative. Vous pouvez indiquer de masquer les relations dont le niveau d'importance est supérieur à une valeur spécifiée.

Qualité de l'ajustement : statistiques et valeurs extrêmes

Tables de statistiques d'ajustement du modèle et des valeurs extrêmes pour les séries cibles sélectionnées pour leur affichage. Ces tableaux contiennent les mêmes informations que les tableaux de la visualisation de système de modèle global. Ces tableaux prennent en charge toutes les fonctions standard pour pivotement et édition de tableaux.

Effets du modèle et paramètres du modèle

Tables des tests d'effets du modèle et des paramètres du modèle pour les séries cibles sélectionnées pour leur affichage. Les tests des effets de modèle comprennent les statistiques F et la valeur de signification associée pour chaque entrée incluse dans le modèle.

Diagramme d'impact

Affiche une représentation graphique des relations de causalité entre une série d'intérêt et d'autres séries qui sont affectées par elle ou qui l'affectent. Les séries qui affectent les séries d'intérêt sont appelées *causes*. La sélection de **Effets** génère un diagramme d'impact qui est initialisé pour afficher les effets. La sélection de **Causes** génère un diagramme d'impact qui est initialisé pour afficher les causes. La sélection de **Causes et effets** génère deux diagrammes d'impact distincts, l'un étant initialisé pour afficher les causes, et l'autre, pour afficher les effets. Vous pouvez basculer interactivement entre causes et effets dans l'élément en sortie qui affiche le diagramme d'impact.

Vous pouvez sélectionner le nombre de niveaux de causes ou d'effets à afficher, le premier niveau étant simplement la série qui vous intéresse. Chaque niveau supplémentaire présente des causes ou des effets plus indirects de la série qui vous intéresse. Par exemple, le troisième niveau dans l'affichage consiste des séries contenant des séries dans le second niveau en tant qu'entrée directe. Les séries du troisième niveau sont donc affectées indirectement par la série qui vous intéresse puisque celle-ci est une entrée directe pour les séries du second niveau.

Tracé des séries

Tracé des valeurs observées et des valeurs prévues pour les séries cibles sélectionnées pour leur affichage. Lorsque des prévisions sont demandées, le tracé indique également les valeurs anticipées et les intervalles de confiance pour les prévisions.

Tracé des résiduels

Tracés des résidus de modèle pour les séries cibles qui sont sélectionnées pour affichage.

Entrées principales

Tracé de chaque cible affichée, au fil du temps, avec les trois principales entrées pour la cible. Il s'agit des entrées avec la plus faible valeur de signification. Pour adapter différentes échelles aux entrées et à la cible, l'axe des y représente le score z pour chaque série.

Tableau des prévisions

Tableaux contenant les valeurs prédites et les intervalles de confiance de ces prévisions pour les séries cible qui sont sélectionnées pour affichage.

Analyse de cause première de valeur extrême

Détermine les séries qui constituent la cause la plus probable de chaque valeur extrême dans une série d'intérêt. Une analyse de la cause première de valeur extrême est effectuée pour chaque série cible incluse dans la liste des séries individuelles dans les paramètres **Séries à afficher**.

Sortie

Tableau et graphique interactifs des valeurs extrêmes

Tableau et graphique des valeurs extrêmes et des causes premières de celles-ci pour chaque série d'intérêt. Le tableau contient une ligne unique pour chaque valeur extrême. Le graphique est un diagramme d'impact. Lorsque vous sélectionnez une ligne du tableau, dans le diagramme d'impact, le chemin entre la série d'intérêt et la série qui constitue la cause la plus probable de la valeur extrême associée est mis en évidence.

Tableau croisé des valeurs extrêmes

Tableau des valeurs extrêmes et des causes premières de ces celles-ci pour chaque série d'intérêt. Cette table contient les mêmes informations que la table dans l'affichage interactif. Elle prend en charge toutes les fonctions standard pour pivotement et édition de tables.

Niveaux de causalité

Vous pouvez spécifier le nombre de niveaux à inclure dans la recherche des causes premières. Le concept de niveaux utilisé ici est le même que celui décrit pour les diagrammes d'impact.

Qualité du modèle par rapport à tous les modèles

Histogramme de la qualité de l'ajustement pour tous les modèles et des statistiques d'ajustement sélectionnées. Les statistiques d'ajustement suivantes sont disponibles :

R-deux

Mesure de qualité d'ajustement d'un modèle linéaire, parfois appelée coefficient de détermination. Il s'agit de la proportion de la variation de la variable cible, expliquée par le modèle. Cela va de 0 à 1. Les petites valeurs indiquent que le modèle ne correspond pas bien aux données.

Pourcentage d'erreur quadratique moyenne

Mesure illustrant l'ampleur de l'écart entre les valeurs prévues par le modèle et les valeurs observées dans les séries. Elle est indépendante des unités utilisées et peut donc servir pour comparer les séries comportant des unités différentes.

Erreur quadratique moyenne.

La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la séries dépendante.

code identificateur de banque

Critère d'information bayésien. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. Le critère BIC (Bayesian Information Criterion) "pénalise" lui aussi les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais de façon plus stricte que le critère AIC (Akaike's Information Criterion).

AIC

Critère d'information d'Akaike. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. L'AIC "pénalise" les modèles surparamétrés (modèles complexes avec un grand nombre d'entrées, par exemple).

Valeurs extrêmes dans le temps

Graphique à barres du nombre de valeurs extrêmes sur toutes les cibles pour chaque intervalle de temps au cours de la période d'estimation.

Transformations de séries

Tableau des transformations éventuelles appliquées aux séries dans le système modèle. Les transformations possibles sont les valeurs d'imputation, d'agrégation et de distribution manquantes.

Période d'estimation

Par défaut, la période d'estimation débute à l'heure de l'observation la plus ancienne et se termine à l'heure de l'observation la plus récente dans toutes les séries.

Par heure de début et de fin

Vous pouvez spécifier le début et la fin de la période d'estimation ou vous pouvez indiquer uniquement le début ou la fin. Si vous omettez d'indiquer le début ou la fin de la période d'estimation, la valeur par défaut est utilisée.

- Si les observations sont définies par un champ de date/heure, entrez des valeurs de début et de fin dans le même format que celui utilisé pour le champ de date/heure.
- Pour les observations définies par des périodes cycliques, spécifiez une valeur pour chacune des zones de période cyclique. Chaque champ est affiché dans une colonne distincte.

- Si une spécification de date est en vigueur pour l'ensemble de données actif, vous devez spécifier une valeur pour chaque composant (par exemple, le mois) de la spécification de date. Chaque composant est affiché dans une colonne distincte.
- Lorsque les observations sont définies par ordre des enregistrements, le début et la fin de la période d'estimation sont définis par le numéro de ligne (tel qu'affiché dans l'éditeur de données) de l'observation pertinente.

Par intervalle de temps le plus récent ou le plus ancien

Définit la période d'estimation en tant que nombre d'intervalles de temps spécifié qui débutent avec l'intervalle de temps le plus ancien ou se termine avec l'intervalle de temps le plus récent dans les données, avec un décalage facultatif. Dans ce contexte, l'intervalle de temps se réfère à celui de l'analyse. Par exemple, supposons que les observations soient de type mensuel et que l'analyse soit de type trimestriel. Si vous spécifiez **Le plus récent** avec la valeur 24 pour le **Nombre d'intervalles de temps**, ceci signifie que la période d'estimation porte sur les 24 derniers trimestres.

Vous avez la possibilité de spécifier un nombre d'intervalles de temps à exclure. Par exemple, si vous spécifiez les 24 intervalles de temps les plus récents et 1 pour le nombre d'intervalles à exclure, la période d'estimation couvrira les 24 intervalles qui précèdent le plus récent.

Prévision

L'option **Etendre les enregistrements dans le futur** permet de définir le nombre d'intervalles de temps pour la prévision au-delà de la fin de la période d'estimation. Dans ce cas, l'intervalle de temps est celui de l'analyse qui est spécifié sur l'onglet Spécifications de données. Lorsque des prévisions sont demandées, des modèles autorégressifs sont générés automatiquement pour les séries d'entrée qui ne sont pas aussi des cibles. Ces modèles sont ensuite utilisés afin de générer des valeurs pour ces séries d'entrée dans la période de prévision.

Enregistrement

Options de destination

Vous pouvez enregistrer les deux transformations des données (telles que l'agrégation ou l'imputation de valeurs manquantes) et les nouvelles variables (spécifiées dans les paramètres **Enregistrer les cibles**) dans un fichier de données IBM SPSS Statistics ou un nouveau jeu de données dans la session en cours. Les valeurs de date/heure contenues dans les données enregistrées sont alignées sur le début de chaque intervalle de temps, tel que le premier du mois, et représentent l'intervalle de temps de l'analyse pour le système de modèle. Vous pouvez enregistrer n'importe quelle nouvelle variable dans le jeu de données actif uniquement si les observations sont définies par une spécification de date ou un ordre d'enregistrement et que les données ne sont pas agrégées.

Enregistrer les cibles

Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résiduels en tant que nouvelles variables. Chaque cible spécifiée à enregistrer génère son propre ensemble de nouvelles variables, et chaque nouvelle variable contient des valeurs pour les périodes de prévision et d'estimation. Pour les valeurs prédites, les intervalles de confiance et les résidus de bruit, vous pouvez spécifier le nom de racine à utiliser en tant que préfixe pour les nouvelles variables. Le nom de variable complet est la concaténation du nom de racine et du nom du champ qui contient la série cible. Le nom de racine doit être conforme aux règles relatives aux noms de variable. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables.

Indiquer les observations contenant des prévisions

Crée une variable qui indique si un enregistrement contient des données de prévision. Vous pouvez spécifier le nom de variable. La valeur par défaut est *ForecastIndicator*.

Cibles à enregistrer

Indique si les nouvelles variables sont créées pour toutes les séries cible du système de modèle or uniquement pour les séries cible spécifiées dans les paramètres **Séries à afficher**.

Exporter le système de modèle

Enregistre le système de modèle dans une archive de fichier compressé (fichier .zip). Le fichier de système de modèle peut être utilisé par la procédure Prévisions de modèle de causalité temporelle pour obtenir des prévisions mises à jour ou pour générer n'importe quelle sortie disponible. Il peut également être utilisé par la procédure Scénarios de modèle de causalité temporelle pour exécuter des analyses de scénario.

Sortie interactive

La sortie de la modélisation de causalité temporelle comporte un certain nombre d'objets en sortie interactifs. Les fonctions interactives sont disponibles en activant (en cliquant deux fois sur) l'objet dans le visualiseur de sortie.

Système de modèle global

Affiche les relations de causalité entre les séries dans le système modèle. Toutes les lignes qui connectent une cible donnée à ses entrées sont de la même couleur. L'épaisseur de la ligne indique le niveau d'importance de la connexion causale, les lignes plus épaisses dénotant une importance plus significative. Les entrées qui ne constituent pas également des cibles sont indiquées par un carré noir.

- Vous pouvez afficher les relations pour les principaux modèles, une série spécifiée, toutes les séries, ou les modèles dépourvus d'entrées. Les modèles principaux sont ceux conformes aux critères spécifiés pour les modèles avec la meilleure adéquation dans les paramètres **Séries à afficher**.
- Vous pouvez générer des diagrammes d'impact pour une ou plusieurs séries en sélectionnant le nom de ces séries dans le diagramme, en cliquant avec le bouton droit de la souris, puis en sélectionnant **Créer un diagramme d'impact** dans le menu contextuel.
- Vous pouvez choisir de masquer les relations de causalité ayant un niveau d'importance supérieur à la valeur spécifiée. Des niveaux d'importance moindres indiquent une relation de causalité plus significative.
- Vous pouvez afficher les relations d'une série donnée en sélectionnant son nom dans le graphique, en cliquant avec le bouton droit de la souris et en sélectionnant **Mettre en évidence les relations des séries** dans le menu contextuel.

Diagramme d'impact

Affiche une représentation graphique des relations de causalité entre une série d'intérêt et d'autres séries qui sont affectées par elle ou qui l'affectent. Les séries qui affectent les séries d'intérêt sont appelées *causes*.

- Vous pouvez changer les séries qui vous intéressent en spécifiant le nom de la série voulue. Un double-clic sur un noeud dans le diagramme d'impact remplace la série qui vous intéresse par celle associée à ce noeud.
- Vous pouvez basculer l'affichage entre causes et effets et modifier le nombre de niveaux de causes et effets à afficher.
- Lorsque vous cliquez une fois sur un noeud, un diagramme de séquence détaillé relatif à la série qui est associée à ce noeud s'ouvre.

Analyse de cause première de valeur extrême

Détermine les séries qui constituent la cause la plus probable de chaque valeur extrême dans une série d'intérêt.

- Vous pouvez afficher la cause première d'une valeur extrême en sélectionnant la ligne correspondant à cette dernière dans le tableau Valeurs extrêmes. Vous pouvez également afficher la cause première en cliquant sur l'icône représentant la valeur extrême dans le graphique séquentiel.
- Lorsque vous cliquez une fois sur un noeud, un diagramme de séquence détaillé relatif à la série qui est associée à ce noeud s'ouvre.

Qualité de modèle global

Histogramme d'adéquation pour tous les modèles et pour une statistique d'adéquation spécifique. Un clic sur une barre du diagramme filtre le tracé des points en n'affichant que les modèles associés à

la barre sélectionnée. Vous pouvez rechercher le modèle d'une série cible spécifique dans le tracé de points en spécifiant le nom de la série.

Distribution des valeurs extrêmes

Graphique à barres du nombre de valeurs extrêmes sur toutes les cibles pour chaque intervalle de temps au cours de la période d'estimation. Un clic sur une barre du diagramme filtre le tracé des points en n'affichant que les valeurs extrêmes associées à la barre sélectionnée.

Remarque :

- Si vous enregistrez un document de sortie contenant une sortie interactive à partir de la modélisation de causalité temporelle et souhaitez conserver les fonctions interactives, vérifiez que l'option **Stocker les informations sur le modèle requises dans le document de sortie** est sélectionnée dans la boîte de dialogue Enregistrer la sortie sous.
- Certaines caractéristiques interactives requièrent que l'ensemble de données actif corresponde aux données utilisées pour la construction du système du modèle de causalité temporelle.

Application de modèles de causalité temporelle

Application de modèles de causalité temporelle

Deux procédures vous permettent d'appliquer des modèles qui ont été créés à l'aide de la procédure Modélisation de causalité temporelle. Les deux procédures nécessitent le fichier de système de modèle qui peut être enregistré dans le cadre de la procédure Modélisation de causalité temporelle.

Prévisions de modèle de causalité temporelle

Cette procédure vous permet d'obtenir des prévisions pour des séries pour lesquelles d'autres données en cours sont disponibles, sans avoir à recréer les modèles. Vous pouvez également générer une sortie disponible à l'aide de la procédure Modélisation de causalité temporelle.

Scénarios de modèle de causalité temporelle

Utilisez cette procédure pour déterminer de quelle manière les valeurs spécifiques d'une série temporelle donnée dans un système de modèle affectent les valeurs prédites des séries temporelles qui ont une relation de causalité avec cette série temporelle.

Prévisions de modèle de causalité temporelle

La procédure Prévision de modèle de causalité temporelle charge un fichier de système de modèle qui a été créé par la procédure Modélisation de causalité temporelle et applique les modèles au jeu de données actif. Cette procédure vous permet d'obtenir des prévisions pour des séries pour lesquelles d'autres données en cours sont disponibles, sans avoir à recréer les modèles. Vous pouvez également générer une sortie disponible à l'aide de la procédure Modélisation de causalité temporelle.

Hypothèses

- La structure des données (basées sur les colonnes ou multidimensionnelles) du jeu de données actif doit être identique à celle qui a été utilisée lorsque le système de modèle a été créé. Pour les données multidimensionnelles, les champs de dimension doivent être les mêmes que ceux utilisés pour créer le système de modèle. En outre, les valeurs de dimension qui ont été utilisées pour créer le système de modèle doivent exister dans le jeu de données actif.
- Les modèles sont appliqués à des champs du jeu de données actif ayant les mêmes noms que les champs indiqués dans le modèle.
- Le ou les champs qui définissaient les observations lorsque le système de modèle a été créé, doivent exister dans le jeu de données actif. On suppose que l'intervalle de temps entre les observations est le même que celui qui existait lorsque tous les modèles ont été créés. Si les observations ont été définies par une spécification de date, la même spécification de date doit exister dans le jeu de données actif. Les spécifications de date sont créées à partir de la boîte de dialogue de définition de dates ou de la commande DATE.

- L'intervalle de temps de l'analyse et les paramètres relatifs à l'agrégation, à la distribution et aux valeurs manquantes sont les mêmes que ceux qui existaient lorsque les modèles ont été créés.

Utilisation des prévisions de modèle de causalité temporelle

Cette fonction nécessite l'option Statistics Forecasting.

À partir des menus, sélectionnez :

Analyser > Prévision > Appliquer les modèles de causalité temporelle ...

1. Entrez la spécification de fichier d'un fichier de système modèle ou cliquez sur **Parcourir** et sélectionnez un fichier de système modèle. Les fichiers de système modèle sont créés avec le procédure Modélisation de causalité temporelle.
2. Cliquez sur l'option permettant d'effectuer une nouvelle estimation des modèles, de créer des prévisions et de générer des sorties.
3. Cliquez sur **Continuer**
4. Indiquez si vous souhaitez utiliser des paramètres de modèle existants ou effectuer une nouvelle estimation des paramètres de modèle à partir des données du jeu de données actif.
5. Spécifiez la portée future des prévisions ou indiquez que vous ne souhaitez pas effectuer de prévisions.
6. Cliquez sur **Options** pour spécifier le contenu de la sortie.
7. Cliquez sur **Enregistrer** pour enregistrer les prédictions, et exportez le système de modèle mis à jour dans un fichier externe lorsque les paramètres de modèle font l'objet d'une nouvelle estimation.
8. Cliquez sur **Exécuter** pour exécuter la procédure.

Paramètres de modèle et prévisions

Charger à partir du fichier de modèle

Les prévisions sont produites à l'aide des paramètres de modèle du fichier de système de modèle et des données du jeu de données actif, sans avoir à effectuer une nouvelle estimation des paramètres de modèle. Les mesures de la qualité de l'ajustement affichées dans la sortie et utilisées pour sélectionner les modèles les plus adaptés sont extraites du fichier de système de modèle. Elles reflètent les données utilisées pour le développement de chaque modèle (ou pour leur dernière mise à jour). L'utilisation de cette option est pertinente pour générer des prévisions et des sorties à partir des données qui ont été utilisées pour construire le système de modèle.

Effectuer une nouvelle estimation à partir des données

Une nouvelle estimation des paramètres de modèle est effectuée à l'aide des données du jeu de données actif. La nouvelle estimation des paramètres de modèle n'affecte pas les entrées qui sont incluses dans le modèle pour chaque cible. L'utilisation de cette option est pertinente lorsque vous disposez de nouvelles données au-delà de la période d'estimation initiale et que vous souhaitez générer des prévisions ou d'autres sorties avec les données mises à jour.

Toutes les observations

Spécifie que la période d'estimation débute à l'heure de l'observation la plus ancienne et se termine à l'heure de l'observation la plus récente dans toutes les séries.

Par heure de début et de fin

Vous pouvez spécifier le début et la fin de la période d'estimation ou vous pouvez indiquer uniquement le début ou la fin. Si vous omettez d'indiquer le début ou la fin de la période d'estimation, la valeur par défaut est utilisée.

- Si les observations sont définies par un champ de date/heure, entrez des valeurs de début et de fin dans le même format que celui utilisé pour le champ de date/heure.
- Pour les observations définies par des périodes cycliques, spécifiez une valeur pour chacune des zones de période cyclique. Chaque champ est affiché dans une colonne distincte.

- Si une spécification de date est en vigueur pour l'ensemble de données actif, vous devez spécifier une valeur pour chaque composant (par exemple, le mois) de la spécification de date. Chaque composant est affiché dans une colonne distincte.
- Lorsque les observations sont définies par ordre des enregistrements, le début et la fin de la période d'estimation sont définis par le numéro de ligne (tel qu'affiché dans l'éditeur de données) de l'observation pertinente.

Par intervalle de temps le plus récent ou le plus ancien

Définit la période d'estimation en tant que nombre d'intervalles de temps spécifié qui débutent avec l'intervalle de temps le plus ancien ou se termine avec l'intervalle de temps le plus récent dans les données, avec un décalage facultatif. Dans ce contexte, l'intervalle de temps se réfère à celui de l'analyse. Par exemple, supposons que les observations soient de type mensuel et que l'analyse soit de type trimestriel. Si vous spécifiez **Le plus récent** avec la valeur 24 pour le **Nombre d'intervalles de temps**, ceci signifie que la période d'estimation porte sur les 24 derniers trimestres.

Vous avez la possibilité de spécifier un nombre d'intervalles de temps à exclure. Par exemple, si vous spécifiez les 24 intervalles de temps les plus récents et 1 pour le nombre d'intervalles à exclure, la période d'estimation couvrira les 24 intervalles qui précèdent le plus récent.

Etendre les enregistrements dans le futur

Définit le nombre d'intervalles de temps pour la prévision au-delà de la fin de la période d'estimation. Dans ce cas, l'intervalle de temps est celui de l'analyse. Lorsque des prévisions sont demandées, des modèles autorégressifs sont générés automatiquement pour les séries d'entrée qui ne sont pas aussi des cibles. Ces modèles sont ensuite utilisés afin de générer des valeurs pour ces séries d'entrée dans la période de prévision afin d'obtenir les prévisions pour les cibles de ces entrées.

Options générales

Largeur de l'intervalle de confiance (%)

Ce paramètre contrôle les intervalles de confiance pour les prévisions et pour les paramètres du modèle. Vous pouvez spécifier n'importe quelle valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Seuil valeurs extrêmes (%)

Une observation est marquée comme étant une valeur extrême si la probabilité, calculée depuis le modèle, qu'il s'agisse d'une valeur extrême dépasse ce seuil. Vous pouvez spécifier une valeur comprise entre 50 et 100.

Séries à afficher

Ces options spécifient les séries (cibles ou entrées) pour lesquelles une sortie est affichée. Le contenu de la sortie pour les séries spécifiées est déterminé par les paramètres **Options de sortie**.

Afficher les cibles associées aux modèles les plus adaptés

Par défaut, la sortie est affichée pour les cibles qui sont associées aux 10 modèles les plus adaptés, comme l'indique la valeur R-deux. Vous pouvez spécifier un nombre fixe différent ou un pourcentage pour les modèles les plus adaptés. Vous pouvez également sélectionner l'une des mesures suivantes d'adéquation des modèles :

R-deux

Mesure de qualité d'ajustement d'un modèle linéaire, parfois appelée coefficient de détermination. Il s'agit de la proportion de la variation de la variable cible, expliquée par le modèle. Cela va de 0 à 1. Les petites valeurs indiquent que le modèle ne correspond pas bien aux données.

Pourcentage d'erreur quadratique moyenne

Mesure illustrant l'ampleur de l'écart entre les valeurs prévues par le modèle et les valeurs observées dans les séries. Elle est indépendante des unités utilisées et peut donc servir pour comparer les séries comportant des unités différentes.

Erreur quadratique moyenne.

La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la séries dépendante.

code identificateur de banque

Critère d'information bayésien. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. Le critère BIC (Bayesian Information Criterion) "pénalise" lui aussi les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais de façon plus stricte que le critère AIC (Akaike's Information Criterion).

AIC

Critère d'information d'Akaike. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. L'AIC "pénalise" les modèles surparamétrés (modèles complexes avec un grand nombre d'entrées, par exemple).

Spécifier des séries individuelles

Vous pouvez spécifier des séries individuelles pour lesquelles générer une sortie.

- Pour les données basées colonnes, spécifiez les champs contenant les séries voulues. L'ordre des champs spécifiés détermine leur ordre d'apparition dans la sortie.
- Pour les données multidimensionnelles; spécifiez une série spécifique en ajoutant une entrée dans la grille pour le champ de métrique qui contient la série. Indiquez ensuite les valeurs des champs de dimension qui définissent la série.
 - Vous pouvez entrer la valeur de chaque champ de dimension directement dans la grille ou vous pouvez effectuer une sélection dans la liste de valeurs de dimension disponibles. Pour effectuer une sélection dans cette liste, cliquez sur les points de suspension dans la cellule correspondant à la dimension voulue. Cette action ouvre la sous-boîte de dialogue Sélectionner des valeurs de dimension.
 - Vous pouvez effectuer une recherche dans la liste des valeurs de dimension depuis la sous-boîte de dialogue Sélectionnez une valeur pour Dimension en cliquant sur l'icône en forme de jumelles et en spécifiant un terme de recherche. Les espaces sont considérés faire partie du terme de recherche. Les astérisques (*) utilisés dans le terme recherché ne représentent pas des caractères génériques.
 - L'ordre des séries dans la grille détermine leur ordre d'apparition dans la sortie.

Pour les données basées sur les colonnes et les données multidimensionnelles, la sortie est limitée à 30 séries. Cette limite inclut les séries individuelles (entrées ou cibles) que vous spécifiez, ainsi que les cibles associées aux modèles les plus adaptés. Les séries spécifiées individuellement sont prioritaires sur les cibles associées aux modèles les plus adaptés.

Options de sortie

Ces options spécifient le contenu de la sortie. Les options du groupe **Sortie pour cibles** génèrent une sortie pour les cibles associées avec les modèles les plus adaptés dans les paramètres **Séries à afficher**. Les options du groupe **Sortie pour séries** génèrent une sortie pour les séries individuelles spécifiées dans les paramètres **Séries à afficher**.

Système de modèle global

Affiche une représentation graphique des relations de causalité entre les séries dans le système modèle. Les tableaux contenant les statistiques et les valeurs extrêmes de la qualité de l'ajustement pour les cibles affichées sont inclus dans le cadre de l'élément de sortie. Lorsque cette option est sélectionnée dans le groupe **Sortie pour les séries**, un élément de sortie distinct est créé pour chaque série individuelle qui est spécifiée dans les paramètres **Séries à afficher**.

Les relations de causalité ont un niveau d'importance associé, où un niveau plus faible indique une connexion plus significative. Vous pouvez indiquer de masquer les relations dont le niveau d'importance est supérieur à une valeur spécifiée.

Qualité de l'ajustement : statistiques et valeurs extrêmes

Tables de statistiques d'ajustement du modèle et des valeurs extrêmes pour les séries cibles sélectionnées pour leur affichage. Ces tableaux contiennent les mêmes informations que les tableaux de la visualisation de système de modèle global. Ces tableaux prennent en charge toutes les fonctions standard pour pivotement et édition de tableaux.

Effets du modèle et paramètres du modèle

Tables des tests d'effets du modèle et des paramètres du modèle pour les séries cibles sélectionnées pour leur affichage. Les tests des effets de modèle comprennent les statistiques F et la valeur de signification associée pour chaque entrée incluse dans le modèle.

Diagramme d'impact

Affiche une représentation graphique des relations de causalité entre une série d'intérêt et d'autres séries qui sont affectées par elle ou qui l'affectent. Les séries qui affectent les séries d'intérêt sont appelées *causes*. La sélection de **Effets** génère un diagramme d'impact qui est initialisé pour afficher les effets. La sélection de **Causes** génère un diagramme d'impact qui est initialisé pour afficher les causes. La sélection de **Causes et effets** génère deux diagrammes d'impact distincts, l'un étant initialisé pour afficher les causes, et l'autre, pour afficher les effets. Vous pouvez basculer interactivement entre causes et effets dans l'élément en sortie qui affiche le diagramme d'impact.

Vous pouvez sélectionner le nombre de niveaux de causes ou d'effets à afficher, le premier niveau étant simplement la série qui vous intéresse. Chaque niveau supplémentaire présente des causes ou des effets plus indirects de la série qui vous intéresse. Par exemple, le troisième niveau dans l'affichage consiste des séries contenant des séries dans le second niveau en tant qu'entrée directe. Les séries du troisième niveau sont donc affectées indirectement par la série qui vous intéresse puisque celle-ci est une entrée directe pour les séries du second niveau.

Tracé des séries

Tracé des valeurs observées et des valeurs prévues pour les séries cibles sélectionnées pour leur affichage. Lorsque des prévisions sont demandées, le tracé indique également les valeurs anticipées et les intervalles de confiance pour les prévisions.

Tracé des résiduels

Tracés des résidus de modèle pour les séries cibles qui sont sélectionnées pour affichage.

Entrées principales

Tracé de chaque cible affichée, au fil du temps, avec les trois principales entrées pour la cible. Il s'agit des entrées avec la plus faible valeur de signification. Pour adapter différentes échelles aux entrées et à la cible, l'axe des y représente le score z pour chaque série.

Tableau des prévisions

Tableaux contenant les valeurs prédites et les intervalles de confiance de ces prévisions pour les séries cible qui sont sélectionnées pour affichage.

Analyse de cause première de valeur extrême

Détermine les séries qui constituent la cause la plus probable de chaque valeur extrême dans une série d'intérêt. Une analyse de la cause première de valeur extrême est effectuée pour chaque série cible incluse dans la liste des séries individuelles dans les paramètres **Séries à afficher**.

Sortie

Tableau et graphique interactifs des valeurs extrêmes

Tableau et graphique des valeurs extrêmes et des causes premières de celles-ci pour chaque série d'intérêt. Le tableau contient une ligne unique pour chaque valeur extrême. Le graphique est un diagramme d'impact. Lorsque vous sélectionnez une ligne du tableau, dans le diagramme d'impact, le chemin entre la série d'intérêt et la série qui constitue la cause la plus probable de la valeur extrême associée est mis en évidence.

Tableau croisé des valeurs extrêmes

Tableau des valeurs extrêmes et des causes premières de ces celles-ci pour chaque série d'intérêt. Cette table contient les mêmes informations que la table dans l'affichage interactif. Elle prend en charge toutes les fonctions standard pour pivotement et édition de tables.

Niveaux de causalité

Vous pouvez spécifier le nombre de niveaux à inclure dans la recherche des causes premières. Le concept de niveaux utilisé ici est le même que celui décrit pour les diagrammes d'impact.

Qualité du modèle par rapport à tous les modèles

Histogramme de la qualité de l'ajustement pour tous les modèles et des statistiques d'ajustement sélectionnées. Les statistiques d'ajustement suivantes sont disponibles :

R-deux

Mesure de qualité d'ajustement d'un modèle linéaire, parfois appelée coefficient de détermination. Il s'agit de la proportion de la variation de la variable cible, expliquée par le modèle. Cela va de 0 à 1. Les petites valeurs indiquent que le modèle ne correspond pas bien aux données.

Pourcentage d'erreur quadratique moyenne

Mesure illustrant l'ampleur de l'écart entre les valeurs prévues par le modèle et les valeurs observées dans les séries. Elle est indépendante des unités utilisées et peut donc servir pour comparer les séries comportant des unités différentes.

Erreur quadratique moyenne.

La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la séries dépendante.

code identificateur de banque

Critère d'information bayésien. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. Le critère BIC (Bayesian Information Criterion) "pénalise" lui aussi les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais de façon plus stricte que le critère AIC (Akaike's Information Criterion).

AIC

Critère d'information d'Akaike. Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance réduit -2. Les petites valeurs indiquent de meilleurs modèles. L'AIC "pénalise" les modèles surparamétrés (modèles complexes avec un grand nombre d'entrées, par exemple).

Valeurs extrêmes dans le temps

Graphique à barres du nombre de valeurs extrêmes sur toutes les cibles pour chaque intervalle de temps au cours de la période d'estimation.

Transformations de séries

Tableau des transformations éventuelles appliquées aux séries dans le système modèle. Les transformations possibles sont les valeurs d'imputation, d'agrégation et de distribution manquantes.

Enregistrer

Enregistrer les cibles

Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résiduels en tant que nouvelles variables. Chaque cible spécifiée à enregistrer génère son propre ensemble de nouvelles variables, et chaque nouvelle variable contient des valeurs pour les périodes de prévision et d'estimation. Pour les valeurs prédites, les intervalles de confiance et les résidus de bruit, vous pouvez spécifier le nom de racine à utiliser en tant que préfixe pour les nouvelles variables. Le nom de variable complet est la concaténation du nom de racine et du nom du champ qui contient la série cible. Le nom de racine doit être conforme aux règles relatives aux noms de variable. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables.

Indiquer les observations contenant des prévisions

Crée une variable qui indique si un enregistrement contient des données de prévision. Vous pouvez spécifier le nom de variable. La valeur par défaut est *ForecastIndicator*.

Cibles à enregistrer

Indique si les nouvelles variables sont créées pour toutes les séries cible du système de modèle or uniquement pour les séries cible spécifiées dans les paramètres **Séries à afficher**.

Options de destination

Vous pouvez enregistrer les deux transformations des données (telles que l'agrégation ou l'imputation de valeurs manquantes) et les nouvelles variables (spécifiées dans les paramètres **Enregistrer les cibles**) dans un fichier de données IBM SPSS Statistics ou un nouveau jeu de données dans la session en cours. Les valeurs de date/heure contenues dans les données enregistrées sont alignées sur le début de chaque intervalle de temps, tel que le premier du mois, et représentent l'intervalle de temps de l'analyse pour le système de modèle. Vous pouvez enregistrer n'importe quelle nouvelle variable dans le jeu de données actif uniquement si les observations sont définies par une spécification de date ou un ordre d'enregistrement et que les données ne sont pas agrégées.

Exporter le système de modèle

Enregistre le système de modèle dans une archive de fichier compressé (fichier .zip). Le fichier de système de modèle peut être réutilisé par cette procédure. Il peut également être utilisé par les scénarios de modèle de causalité temporelle pour exécuter des analyses de scénario. Cette option est disponible uniquement lorsque les paramètres de modèle font l'objet d'une nouvelle estimation.

Scénarios de modèle de causalité temporelle

La procédure Scénarios de modèle de causalité temporelle exécute des scénarios définis par l'utilisateur pour un système de modèle causal temporel à partir des données de l'ensemble de données actif. Un *scénario* est défini par une série temporelle, appelée *série racine*, et un ensemble de valeurs définies par l'utilisateur pour cette série sur un intervalle spécifié. Les valeurs spécifiées sont ensuite utilisées pour générer des prévisions pour les séries temporelles qui sont affectées par la série racine. La procédure requiert un fichier de système modèle créé par la procédure Modélisation de causalité temporelle. Il est supposé que l'ensemble de données actif correspond aux mêmes données que celles utilisées pour créer le fichier de système de modèle.

Exemple

A l'aide de la procédure Modélisation de causalité temporelle, un décideur a découvert une métrique clé qui affecte un certain nombre d'indicateurs de performance importants. La métrique étant contrôlable, le décideur désire examiner l'effet de diverses valeurs sur la métrique au cours du trimestre suivant. Pour cela, il suffit de charger le fichier de système de modèle dans la procédure Scénarios de modèle de causalité temporelle et de spécifier les différents jeux de valeurs pour la métrique clé.

Exécution de scénarios de modèle de causalité temporelle

Cette fonction nécessite l'option Statistics Forecasting.

À partir des menus, sélectionnez :

Analyser > Prévision > Appliquer les modèles de causalité temporelle...

1. Entrez la spécification de fichier d'un fichier de système modèle ou cliquez sur **Parcourir** et sélectionnez un fichier de système modèle. Les fichiers de système modèle sont créés avec le procédure Modélisation de causalité temporelle.
2. Cliquez sur l'option pour exécuter des scénarios.
3. Cliquez sur **Continuer**
4. Dans l'onglet Scénarios (dans la boîte de dialogue Scénarios de modèle de causalité temporelle), cliquez sur **Définir la période du scénario** et spécifiez la période du scénario.

5. Pour des données basées colonnes, cliquez sur **Ajouter un scénario** pour définir chaque scénario. Pour des données multidimensionnelles, cliquez sur **Ajouter un scénario** pour définir chaque scénario individuel et sur **Ajouter un groupe de scénarios** pour définir chaque groupe de scénarios.
6. Cliquez sur **Options** pour spécifier le contenu de la sortie et pour indiquer le niveau d'affectation des séries par le scénario.
7. Cliquez sur **Exécuter** pour exécuter la procédure.

Définition de la période de scénario

La période de scénario est la période au cours de laquelle vous spécifiez les valeurs qui sont utilisées pour exécuter vos scénarios. Elle peut débuter avant ou après la fin de la période d'estimation. Vous pouvez éventuellement spécifier que la prévision doit s'étendre au-delà de la fin de la période de scénario. Par défaut, les prévisions sont générées jusqu'à la fin de la période de scénario. Tous les scénarios utilisent la même période de scénario et les mêmes spécifications pour indiquer l'étendue de la prévision.

Remarque : Les prévisions commencent lors du premier intervalle de temps et s'étendent au-delà du début de la période de scénario. Par exemple, si la période de scénario débute le 01/11/2014 et l'intervalle de temps est exprimé en mois, la première prévision concerne le 01/12/2014.

Spécifier par début, fin et nombre de prévisions

- Si les observations sont définies par un champ de date/heure, entrez les valeurs de début, de fin et de prévision au fil du temps en utilisant le même format que celui du champ de date/heure. Les valeurs des champs de date/heure sont alignées sur le début de l'intervalle de temps associé. Par exemple, si l'intervalle de temps de l'analyse est exprimé en mois (Months), la valeur 10/10/2014 est ajustée à 10/01/2014, ce qui représente le début du mois.
- Pour les observations qui sont définies par des périodes cycliques, spécifiez une valeur pour chaque champ de période cyclique. Chaque zone apparaît dans une colonne distincte.
- Si une spécification de date est en vigueur pour le jeu de données actif, vous devez spécifier une valeur pour chaque composant (par exemple, Month) de la spécification de date. Chaque composant apparaît dans une colonne distincte.
- Lorsque les observations sont définies par ordre d'enregistrement, le début, la fin et la prévision au fil du temps sont définis par le numéro de ligne (tel qu'il est affiché dans l'éditeur de données) de l'observation concernée.

Spécifier par intervalles de temps relatifs à la fin de la période d'estimation

Définit le début et la fin pour le nombre d'intervalles de temps relatifs à la fin de la période d'estimation, où l'intervalle de temps est celui de l'analyse. La fin de la période d'estimation est définie comme intervalle de temps 0. Les intervalles de temps avant la fin de la période d'estimation ont des valeurs négatives et des intervalles après la fin de la période d'estimation ont des valeurs positives. Vous pouvez également spécifier un nombre d'intervalles pour l'extension de la prévision au-delà de la fin de la période de scénario. La valeur par défaut est 0.

Supposons par exemple que l'intervalle de temps de l'analyse est exprimé en mois (Months) et que vous indiquez 1 pour l'intervalle de début, 3 pour l'intervalle de fin et 1 pour l'extension de la prévision. La période de scénario correspond alors aux 3 mois qui suivent la fin de la période d'estimation. Les prévisions sont générées pour le deuxième et le troisième mois de la période de scénario et pour 1 mois supplémentaire après la fin de la période de scénario.

Ajout de scénarios et de groupes de scénarios

L'onglet Scénarios spécifie les scénarios à exécuter. Pour définir des scénarios, vous devez d'abord spécifier la période du scénario en cliquant sur **Définir la période du scénario**. Les scénarios et les groupes de scénarios (s'appliquent uniquement aux données multidimensionnelles) sont créés en cliquant sur le bouton **Ajouter un scénario** ou **Ajouter un groupe de scénarios** associé. En sélectionnant un scénario ou un groupe de scénarios spécifique dans la grille associée, vous pouvez le modifier, en faire une copie, ou le supprimer.

Données par colonnes

La colonne **Zone racine** dans la grille spécifie la zone de série temporelle dont remplacer les valeurs par celles du scénario. La colonne **Valeurs de scénario** affiche les valeurs de scénario spécifiées de la plus ancienne à la plus récente. Si les valeurs du scénario sont définies par une expression, la colonne affiche l'expression concernée.

Données multidimensionnelles

Scénarios individuels

Chaque ligne de la grille Scénarios individuels spécifie une série temporelle dont les valeurs sont remplacées par les valeurs de scénario spécifiées. La série est définie par la combinaison de la zone spécifiée dans la colonne **Métrique racine** et de la valeur spécifiée pour chacune des zones de dimension. Le contenu de la colonne **Valeurs de scénario** est identique à celui des données basées sur les colonnes.

Groupes de scénarios

Un *groupe de scénarios* définit un ensemble de scénarios basés sur un champ de métrique racine unique et plusieurs jeux de valeurs de dimension. Chaque ensemble de valeurs de dimension (une valeur par champ de dimension), pour le champ de métrique spécifié, définit une série temporelle. Un scénario individuel est ensuite généré pour chaque série temporelle dont les valeurs sont ensuite remplacées par les valeurs de scénario. Les valeurs de scénario d'un groupe de scénarios sont spécifiées par une expression, laquelle est ensuite appliquée à chaque série temporelle du groupe.

La colonne **Comptage des séries** affiche le nombre d'ensembles de valeurs de dimension qui sont associés à un groupe de scénarios. La valeur affichée peut être supérieure au nombre réel de séries temporelles associées au groupe de scénarios (une série par jeu). Cette condition survient lorsque certaines combinaisons de valeurs de dimension spécifiées ne correspondent pas aux séries contenues par la métrique racine pour le groupe.

Supposons par exemple un groupe de scénarios ayant un champ de métrique *advertising* et deux champs de dimension *region* et *brand*. Vous pouvez définir un groupe de scénarios basé sur *advertising* comme métrique racine et qui inclut toutes les combinaisons de *region* et de *brand*. Vous pourriez alors spécifier *advertising*1.2* comme expression afin d'étudier l'effet d'une augmentation de 20 % de *advertising* pour chacune des séries temporelles associées à la zone *advertising*. Si 4 valeurs de *region* et 2 valeurs de *brand* existent, 8 de ces séries temporelles seront alors présentes, et donc 8 scénarios seront définis par le groupe.

Définition de scénario

Les paramètres de définition d'un scénario varient selon que vous disposez de données basées sur les colonnes ou multidimensionnelles.

Séries racine

Spécifie la série racine pour le scénario. Chaque scénario est basé sur une série racine unique. Pour les données basées sur les colonnes, vous sélectionnez le champ qui définit la série racine. Pour les données multidimensionnelles, spécifiez la série racine en ajoutant une entrée à la grille pour le champ de métrique contenant la série. Spécifiez ensuite les valeurs des champs de dimension qui définissent la série racine. Les considérations suivantes s'appliquent lors de la spécification des valeurs de dimension :

- Vous pouvez entrer la valeur de chaque champ de dimension directement dans la grille ou vous pouvez effectuer une sélection dans la liste de valeurs de dimension disponibles. Pour effectuer une sélection dans cette liste, cliquez sur les points de suspension dans la cellule correspondant à la dimension voulue. Cette action ouvre la sous-boîte de dialogue Sélectionner des valeurs de dimension.
- Vous pouvez effectuer une recherche dans la liste des valeurs de dimension depuis la sous-boîte de dialogue Sélectionner une valeur pour Dimension en cliquant sur l'icône en forme de jumelles et en spécifiant un terme de recherche. Les espaces sont considérés faire partie du terme de recherche. Les astérisques (*) utilisés dans le terme recherché ne représentent pas des caractères génériques.

Spécifier les cibles affectées

Utilisez cette option lorsque vous connaissez les cibles spécifiques qui sont affectées par la série racine et que vous souhaitez déterminer uniquement les effets sur ces cibles. Par défaut, les cibles qui sont affectées par la série racine sont déterminées automatiquement. Vous pouvez spécifier le niveau des séries qui sont affectées par le scénario à l'aide des paramètres de l'onglet Options.

Pour les données basées sur les colonnes, sélectionnez les cibles souhaitées. Pour les données multidimensionnelles, spécifiez une série cible en ajoutant une entrée à la grille du champ de métrique cible qui contient la série. Par défaut, toutes les séries contenues dans le champ de métrique spécifié sont incluses. Vous pouvez personnaliser l'ensemble de séries incluses en personnalisant les valeurs incluses pour un ou plusieurs champs de dimension. Pour personnaliser les valeurs de dimension qui sont incluses, cliquez sur les points de suspension correspondant à la dimension souhaitée. Cette action ouvre la boîte de dialogue Sélectionnez des valeurs pour Dimension.

La colonne **Comptage des séries** (pour les données multidimensionnelles) affiche le nombre d'ensembles de valeurs de dimension actuellement spécifiés pour la métrique cible associée. La valeur affichée peut être supérieure au nombre réel de séries cible affectées (une série par ensemble). Cette condition se produit lorsque certaines des combinaisons de valeurs de dimension spécifiées ne correspondent pas aux séries contenues par la métrique cible associée.

ID du scénario

Chaque scénario doit avoir un identificateur unique. L'identificateur figure dans la sortie qui est associée au scénario. Aucune restriction, hormis l'unicité, ne s'applique à la valeur de l'identificateur.

Spécifier des valeurs de scénario pour les séries racine

Utilisez cette option afin de spécifier des valeurs explicites pour les séries racine dans la période de scénario. Vous devez spécifier une valeur numérique pour chaque intervalle de temps répertorié dans la grille. Vous pouvez obtenir les valeurs de la série racine (réelle ou prévue) pour chaque intervalle dans le scénario en cliquant sur **Lire, Prévision**, ou sur **Lire\Prévision**.

Spécifier une expression pour les valeurs de scénario des séries racine

Vous pouvez définir une expression pour le calcul des valeurs de la série racine dans la période de scénario. Vous pouvez entrer directement l'expression ou cliquer sur le bouton de calculatrice et créer l'expression à partir du générateur d'expressions de valeurs de scénario.

- L'expression peut contenir n'importe quelle cible ou entrée du système modèle.
- Lorsque la période de scénario s'étend au-delà des données existantes, l'expression s'applique aux valeurs prévues des champs de l'expression.
- Pour les données multidimensionnelles, chaque champ de l'expression spécifie une série temporelle qui est définie par le champ et les valeurs de dimension qui ont été spécifiées pour la métrique racine. Ces sont ces séries temporelles qui sont utilisées pour évaluer l'expression.

A titre d'exemple, supposons que la zone racine est *advertising* et l'expression $advertising * 1.2$. Les valeurs *advertising* utilisées dans le scénario représentent un accroissement de 20 % par rapport aux valeurs existantes.

Remarque : Les scénarios sont créés en cliquant sur **Ajouter un scénario** sur l'onglet Scénarios.

Générateur d'expressions de valeurs de scénario

Utilisez le générateur d'expression de valeurs de scénario afin de créer une expression qui sera utilisée pour calculer des valeurs de scénario pour un scénario unique ou pour un groupe de scénarios. Pour construire une expression, collez les composants dans le champ **Expression** ou saisissez-les directement dans le champ **Expression**.

- L'expression peut contenir n'importe quelle cible ou entrée du système modèle.
- Vous pouvez coller des fonctions en sélectionnant un groupe dans la liste Groupe de fonctions, puis en double-cliquant sur la fonction désirée dans la liste Fonctions (ou sélectionnez la fonction, puis cliquez sur la flèche adjacente à la liste Groupe de fonctions). Entrez tous les paramètres indiqués par un point d'interrogation. Le groupe de fonctions libellé **Tous** répertorie toutes les fonctions disponibles.

Une brève description de la fonction sélectionnée apparaît dans une zone particulière de la boîte de dialogue.

- Les constantes chaîne doivent être présentées entre guillemets.
- Si des valeurs contiennent des chiffres décimaux, utilisez la virgule comme indicateur décimal.

Pour accéder au générateur de formules de valeurs de scénario, cliquez sur le bouton de la calculatrice dans la boîte de dialogue Définition du scénario ou Définition du groupe de scénarios.

Sélection de valeurs de dimension

Pour les données multidimensionnelles, vous pouvez personnaliser les valeurs de dimension qui définissent les cibles affectées par un scénario ou un groupe de scénarios. Vous pouvez également personnaliser les valeurs de dimension qui définissent l'ensemble de séries racine pour un groupe de scénarios.

Toutes les valeurs

Spécifie d'inclure toutes les valeurs du champ de dimension actuel. Il s'agit de l'option par défaut.

Sélectionner des valeurs

Utilisez cette option pour spécifier l'ensemble de valeurs pour le champ de dimension actuel. Vous pouvez filtrer l'ensemble des valeurs dans lesquelles choisir. Les valeurs qui répondent à la condition de filtre s'affichent dans l'onglet **Avec correspondance** tandis que les autres valeurs apparaissent dans l'onglet **Sans correspondance** de la liste **Valeurs non sélectionnées**. L'onglet **Tout** répertorie toutes les valeurs non sélectionnées, quelle que soit la condition de filtre.

- Vous pouvez utiliser des astérisques (*) comme caractères génériques lorsque vous spécifiez un filtre.
- Pour désélectionner le filtre en cours, spécifiez une valeur vide pour le terme recherché sur la boîte de dialogue Filtrer les valeurs affichées.

Pour personnaliser les valeurs de dimension pour les cibles affectées :

1. Dans la boîte de dialogue Définition du scénario ou Définition du groupe de scénarios, sélectionnez la métrique cible pour laquelle vous souhaitez personnaliser les valeurs de dimension.
2. Cliquez sur les points de suspension dans la colonne de la dimension que vous désirez personnaliser.

Pour personnaliser les valeurs de dimension pour la série racine d'un groupe de scénarios :

1. Dans la boîte de dialogue Définition du groupe de scénarios, cliquez sur les points de suspension (dans la grille de la série racine) en regard de la dimension que vous désirez personnaliser.

Définition du groupe de scénarios

Séries racine

Spécifie l'ensemble de séries racine pour le groupe de scénarios. Un scénario individuel est généré pour chaque série temporelle de l'ensemble. Spécifiez la série racine en ajoutant une entrée à la grille pour le champ de métrique contenant la série souhaitée. Spécifiez ensuite les valeurs des champs de dimension qui définissent l'ensemble. Par défaut, toutes les séries contenues dans la zone de métrique racine spécifiée sont incluses. Vous pouvez personnaliser l'ensemble de séries incluses en personnalisant les valeurs incluses pour un ou plusieurs champs de dimension. Pour personnaliser les valeurs de dimension à inclure, cliquez sur les points de suspension en regard d'une dimension. Cette action ouvre la boîte de dialogue Sélectionnez des valeurs pour Dimension.

La colonne **Comptage des séries** affiche le nombre d'ensembles de valeurs de dimension actuellement inclus pour la métrique racine associée. La valeur affichée peut être supérieure au nombre réel de séries racine pour le groupe de scénarios (une série par jeu). Cette condition survient lorsque certaines combinaisons de valeurs de dimension spécifiées ne correspondent pas aux séries contenues par la métrique racine.

Spécifiez les séries cibles affectées

Utilisez cette option lorsque vous connaissez les cibles spécifiques qui sont affectées par l'ensemble de séries racine et que vous souhaitez déterminer uniquement les effets sur ces cibles. Par défaut, les cibles affectées par chaque série racine sont déterminées automatiquement. Vous pouvez spécifier

l'amplitude des séries affectées par chaque scénario individuel via des paramètres de l'onglet Options.

Vous spécifiez la série cible en ajoutant une entrée à la grille pour le champ de métrique contenant la série. Par défaut, toutes les séries contenues dans le champ de métrique spécifié sont incluses. Vous pouvez personnaliser l'ensemble de séries incluses en personnalisant les valeurs incluses pour un ou plusieurs champs de dimension. Pour personnaliser les valeurs de dimension qui sont incluses, cliquez sur les points de suspension correspondant à la dimension souhaitée. Cette action ouvre la boîte de dialogue Sélectionnez des valeurs pour Dimension.

La colonne **Comptage des séries** affiche le nombre d'ensembles de valeurs de dimension actuellement spécifiés pour la métrique cible associée. La valeur affichée peut être supérieure au nombre réel de séries cible affectées (une série par ensemble). Cette condition se produit lorsque certaines des combinaisons de valeurs de dimension spécifiées ne correspondent pas aux séries contenues par la métrique cible associée.

Préfixe de l'ID de scénario

Chaque groupe de scénarios doit avoir un préfixe unique. Ce préfixe est utilisé pour construire un identificateur qui est affiché dans la sortie associée à chaque scénario individuel du groupe de scénarios. L'identificateur d'un scénario individuel est le préfixe, suivi successivement d'un trait de soulignement et de la valeur de chaque champ de dimension qui identifie la série racine. Les valeurs de dimension sont séparées par des traits de soulignement. Aucune restriction n'est imposée quant à la valeur du préfixe, à part son unicité.

Expression pour les valeurs de scénario pour les séries racine

Les valeurs de scénario pour un groupe de scénarios sont spécifiées par une expression, laquelle est ensuite utilisée pour calculer les valeurs de chacune des séries racine dans le groupe. Vous pouvez entrer directement une expression ou cliquer sur le bouton de calculatrice et créer l'expression à partir du générateur d'expressions de valeurs de scénario.

- L'expression peut contenir n'importe quelle cible ou entrée du système modèle.
- Lorsque la période de scénario s'étend au-delà des données existantes, l'expression s'applique aux valeurs prévues des champs de l'expression.
- Pour chaque série racine du groupe, les zones de l'expression spécifient les séries temporelles définies par ces zones et les valeurs de dimension qui définissent la série racine. Ces séries temporelles sont celles utilisées pour évaluer l'expression. Par exemple, si une série racine est définie par `region= ' north ' et brand= ' X ' , les séries temporelles qui sont utilisées dans l'expression sont définies par ces mêmes valeurs de dimension.`

A titre d'exemple, supposons que la zone de métrique racine est *advertising* et qu'existent les deux zones de dimension *region* et *brand*. En outre, imaginons que le groupe de scénarios comprend toutes les combinaisons des valeurs de champs de dimension. Vous pourriez alors spécifier `advertising*1.2` comme expression afin d'étudier l'effet d'une augmentation de 20 % de *advertising* pour chacune des séries temporelles associées à la zone *advertising*.

Remarque : Les groupes de scénarios s'appliquent uniquement aux données multidimensionnelles et sont créés en cliquant sur **Ajouter un groupe de scénarios** sur l'onglet Scénarios.

Options

Niveau maximal pour les cibles affectées

Spécifie le nombre maximal de niveaux de cibles affectées. Chaque niveau successif, jusqu'au maximum de 5, inclut des cibles affectées plus indirectement par la série racine. Spécifiquement, le premier niveau inclut des cibles pour lesquelles la série racine est une entrée directe. Les cibles du deuxième niveau ont des cibles du deuxième niveau comme entrée directe, et ainsi de suite. L'augmentation de la valeur de ce paramètre accroît la complexité du calcul et peut affecter les performances.

Nombre maximal de cibles détectées automatiquement

Indique le nombre maximal de cibles affectées qui sont automatiquement détectées pour chaque série racine. L'augmentation de la valeur de ce paramètre accroît la complexité du calcul et peut affecter les performances.

Diagramme d'impact

Affiche une représentation graphique des relations de causalité entre la série racine pour chaque scénario et les séries cibles qu'elle affecte. Des tableaux des valeurs du scénario et des valeurs prévues pour les cibles affectées sont inclus dans le cadre de l'élément en sortie. Le graphique inclut des tracés des valeurs prévues des cibles affectées. Lorsque vous cliquez sur un noeud du diagramme d'impact, un diagramme de séquence détaillé pour la série associée à ce noeud s'ouvre. Un diagramme d'impact distinct est généré pour chaque scénario.

Tracés de séries

Génère des tracés des séries des valeurs prévues pour chaque cible affectée dans chaque scénario.

Tableaux de prévisions et de scénarios

Tableaux contenant les valeurs prédites et les valeurs de scénario pour chaque scénario. Ces tableaux contiennent les mêmes informations que les tableaux du diagramme d'impact. Ces tableaux prennent en charge toutes les fonctions standard pour pivotement et édition de tableaux.

Inclure les intervalles de confiance dans les tracés et les tableaux

Spécifie si les intervalles de confiance des prévisions de scénario sont inclus dans les tableaux de résultats et le graphique.

Largeur d'intervalle de confiance (%) :

Ce paramètre contrôle les intervalles de confiance pour les prévisions du scénario. Vous pouvez spécifier n'importe quelle valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Mesures de la qualité d'ajustement

Cette section fournit les définitions des mesures de la qualité d'ajustement utilisées pour la modélisation de séries chronologiques.

- *R-carré stationnaire*. Mesure qui compare la partie stationnaire du modèle à un modèle de moyenne simple. Cette mesure est préférable à un R-carré ordinaire lorsqu'il y a une tendance ou un motif saisonnier. Le R-carré stationnaire peut avoir une valeur négative comprise sur la plage de l'infini négatif à 1. Des valeurs négatives indiquent que le modèle considéré est moins bon que le modèle de référence. Les valeurs positives signifient que la sous-considération du modèle est meilleure que le modèle de la ligne de base.
- *R au carré*. Estimation de la proportion de la variation totale de la série expliquée par le modèle. Cette mesure est très utile lorsque la série est stationnaire. Le R-carré peut être négatif avec une fourchette allant de l'infini négatif à 1. Des valeurs négatives indiquent que le modèle considéré est moins bon que le modèle de référence. Les valeurs positives signifient que la sous-considération du modèle est meilleure que le modèle de la ligne de base.
- *RMSE*. Erreur quadratique moyenne. La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la série dépendante.
- *MAPE*. Erreur absolue moyenne en pourcentage. Une mesure de la variation de la série dépendante par rapport au niveau prévu par le modèle. Elle est indépendante des mesures utilisées et peut donc être utilisée pour comparer les séries comportant des unités différentes.
- *MAE*. Erreur absolue moyenne. Mesure la variation de la série par rapport au niveau prévu par le modèle. La MAE est reportée dans les unités de séries d'origine.
- *MaxAPE*. Erreur de pourcentage absolue maximale. La plus grande erreur prévue, exprimée en pourcentage. Cette mesure est utile pour imaginer le pire scénario lors de vos prévisions.
- *MaxAE*. Erreur absolue maximale. La plus grande erreur prévue, exprimée dans les mêmes unités que la série dépendante. Comme MaxAPE, elle est utile pour imaginer le pire scénario lors de vos prévisions. L'erreur Maximum absolue et l'erreur de pourcentage absolue maximum peuvent se produire

à différents moments d'une série, par exemple lorsque l'erreur absolue pour la valeur d'une grande série est légèrement supérieure à l'erreur absolue pour la valeur d'une petite série. Dans ce cas, l'erreur maximum absolue se produira à la valeur de la grande série et l'erreur de pourcentage absolue maximum se produira à la valeur de la petite série.

- *BIC normalisé*. Critère d'information bayésien normalisé. Une mesure générale de l'ajustement global d'un modèle qui essaye de prendre en compte la complexité du modèle. C'est un score basé sur l'erreur quadratique moyenne et qui inclut une pénalité pour le nombre de paramètres du modèle et la longueur de la série. La pénalité supprime l'avantage des modèles disposant de plus de paramètres, rendant les statistiques plus faciles à comparer parmi les différents modèles d'une même série.

Types de valeurs extrêmes

Cette section fournit les définitions des types de valeurs extrêmes utilisées pour la modélisation de séries chronologiques.

- *additive*. Valeur extrême qui affecte une observation unique. Par exemple, une erreur de codification des données peut être identifiée comme une valeur extrême additive.
- *changement de niveau*. Valeur extrême qui remplace toutes les observations par une constante, en commençant à un point particulier de la série. Un changement de politique peut entraîner un changement de niveau.
- *innovation*. Valeur extrême qui agit en tant qu'addition au terme de bruit à un point particulier de la série. Pour les séries stationnaires, une valeur extrême d'innovation affecte plusieurs observations. Pour les séries non stationnaires, elle peut affecter chaque observation à partir d'un point particulier de la série.
- *transitoire*. Valeur extrême dont l'impact décroît de façon exponentielle pour atteindre 0.
- *additif saisonnier*. Valeur extrême qui affecte une observation particulière et toutes les observations suivantes séparées par une ou plusieurs périodes saisonnières. Toutes ces observations sont affectées de la même façon. Une valeur extrême additive saisonnière peut se produire si, au début d'une année donnée, les ventes sont supérieures chaque mois de janvier.
- *tendance locale*. Valeur extrême qui apparaît en début de tendance locale à un point spécifique d'une série.
- *correctif additif*. Groupe d'au moins deux valeurs extrêmes additives consécutives. Sélectionner le type de valeur extrême permet de détecter des valeurs extrêmes additives individuelles ainsi que les groupes.

Guide relatif aux tracés ACF/PACF

Les tracés présentés ici se rapportent à des processus ARIMA réels ou théoriques. Voici quelques règles générales permettant d'identifier le processus :

- Les séries non stationnaires possèdent une fonction d'autocorrélation (ACF) qui reste importante pour au moins une demi-douzaine de retards, plutôt que de diminuer rapidement à 0. Vous devez modifier une série de ce type jusqu'à ce qu'elle soit stationnaire afin que vous puissiez identifier le processus.
- Les processus autorégressifs disposent d'une fonction d'autocorrélation (ACF) qui diminue de manière exponentielle et de projections dans les premiers retards de la fonction d'autocorrélation partielle (PACF). Le nombre de projections indique l'ordre de l'autorégression.
- Les processus moyenne mobile disposent de projections dans les premiers retards de la fonction d'autocorrélation et d'une fonction d'autocorrélation partielle qui diminue de manière exponentielle. Le nombre de projections indique l'ordre de la moyenne mobile.
- Les processus mixtes (ARMA) indiquent généralement des baisses exponentielles dans la fonction d'autocorrélation et dans la fonction d'autocorrélation partielle.

Au stade de l'identification, vous n'avez pas à vous soucier du signe de la fonction ACF ou PACF, ni de la vitesse à laquelle une fonction ACF ou PACF qui diminue de manière exponentielle s'approche de 0. Ils

sont fonction du signe et de la valeur réelle des coefficients AR et MA. Dans certains exemples, une ACF qui baisse de manière exponentielle adopte tour à tour des valeurs positives et négatives.

Les tracés ACF et PACF issus de données réelles ne sont jamais aussi nets que ceux montrés ici. Vous devez apprendre à identifier les informations essentielles de tout tracé. Vérifiez toujours l'ACF et la PACF des résidus, au cas où votre identification serait incorrecte. Tenez compte des points suivants :

- Les processus saisonniers affichent ces motifs au niveau des retards saisonniers (multiples de la période saisonnière).
- Vous avez le droit de traiter des valeurs non significatives comme 0. En d'autres termes, vous pouvez ignorer les valeurs qui se situent dans les intervalles de confiance sur les tracés. Cependant, vous ne devez pas les ignorer si elles poursuivent le motif des valeurs significatives au niveau statistique.
- Une autocorrélation intermittente sera significative au niveau statistique uniquement par hasard. Vous pouvez ignorer une autocorrélation statistiquement significative si elle est isolée, de préférence au niveau d'un retard élevé, et si elle ne se produit pas au niveau d'un retard saisonnier.

Pour plus d'informations sur les tracés ACF et PACF, consultez les textes relatifs à l'analyse ARIMA.

Tableau 3. ARIMA(0,0,1), $q > 0$

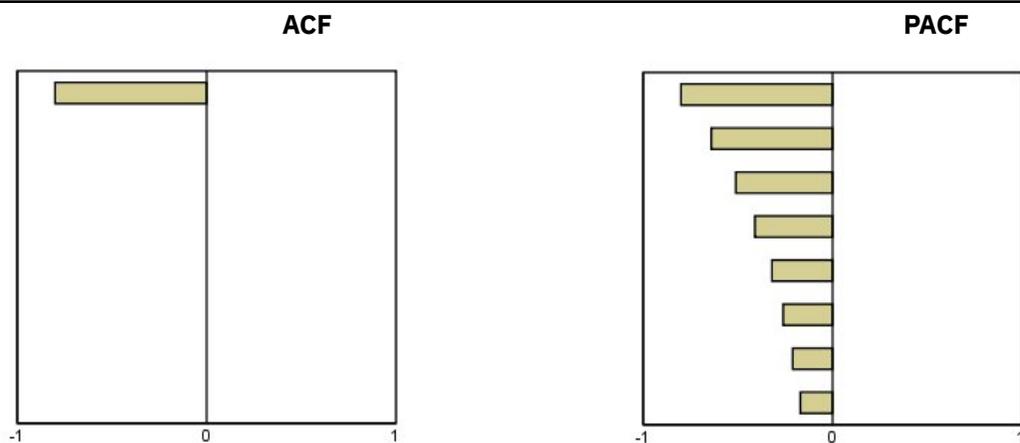
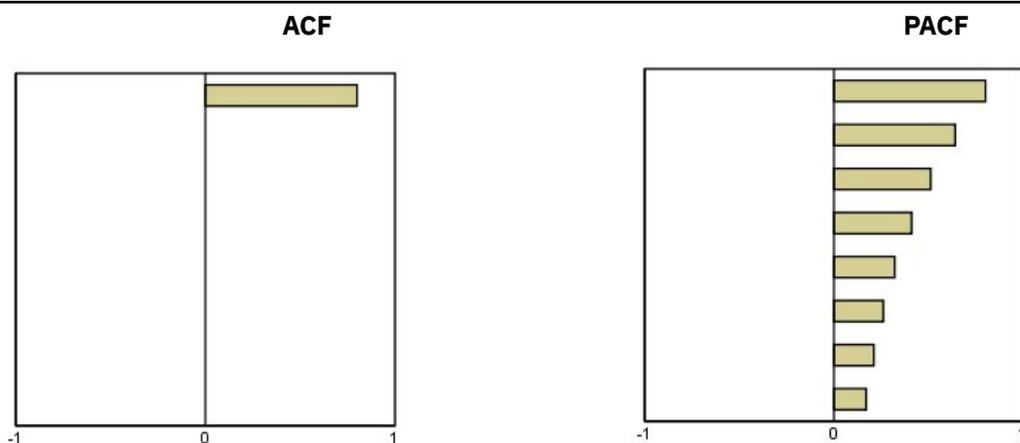


Tableau 4. ARIMA(0,0,1), $q < 0$



ARIMA(0,0,2), $1_2 > 0$

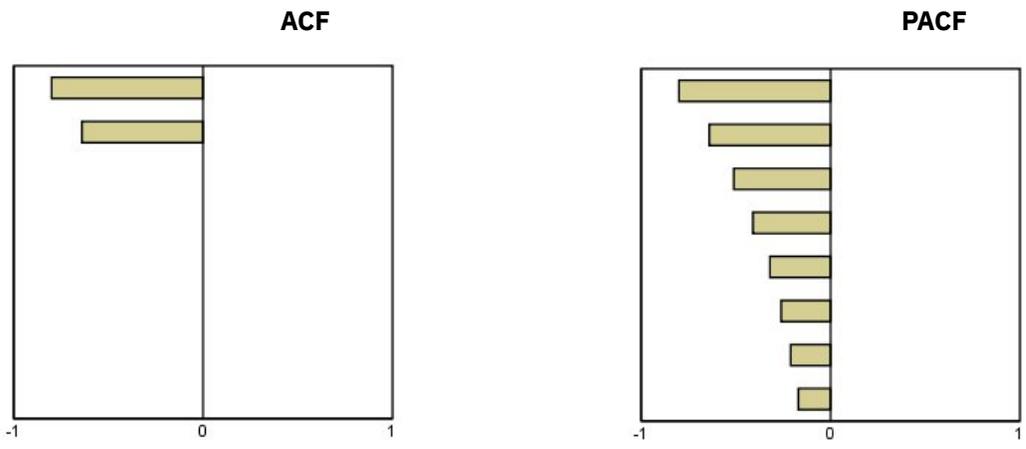


Tableau 5. $ARIMA(1,0,0), f>0$

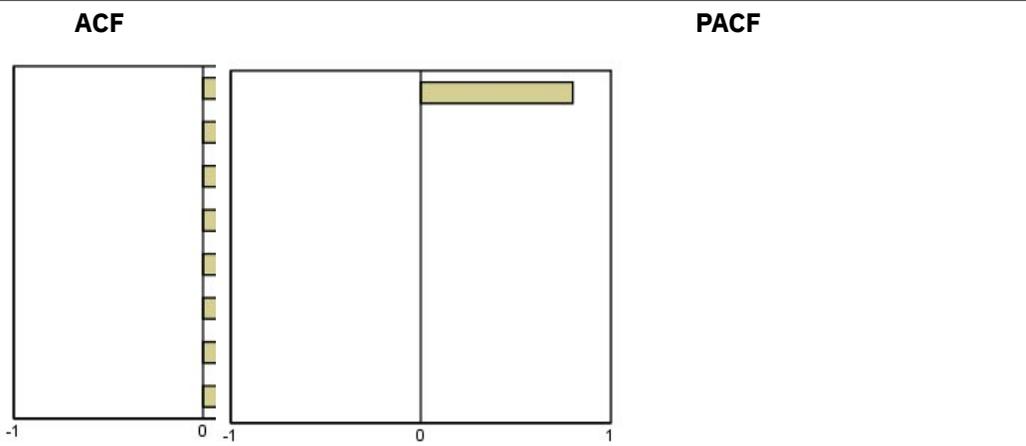
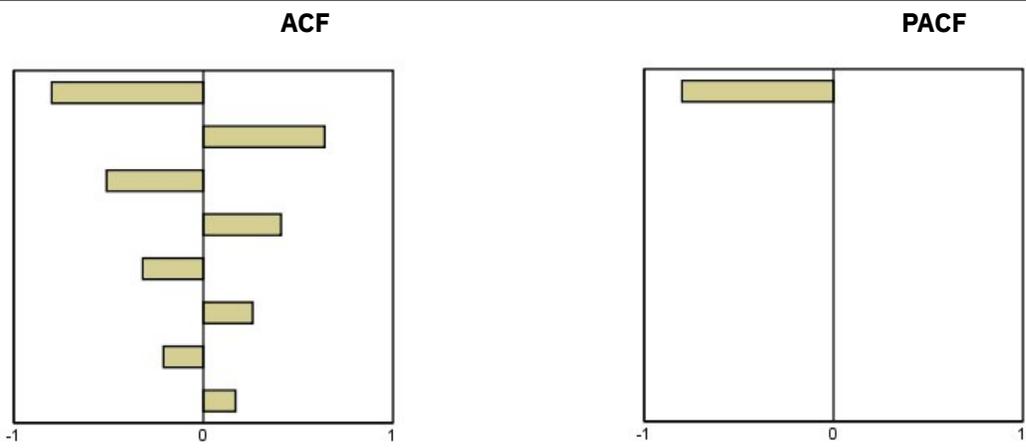
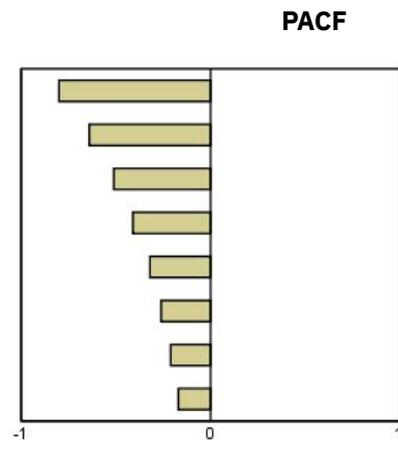
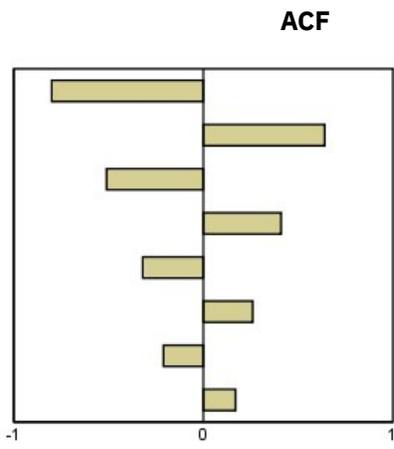


Tableau 6. $ARIMA(1,0,0), f<0$



$ARIMA(1,0,1), <0, >0$



ARIMA(2,0,0), $\phi_1 > 0$

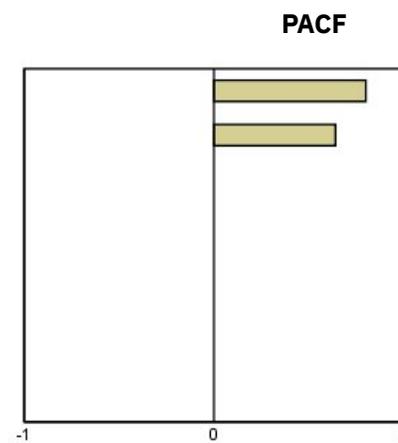
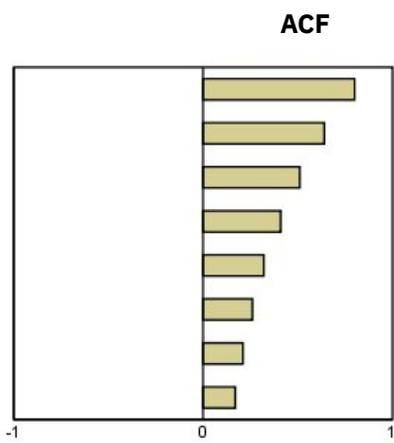
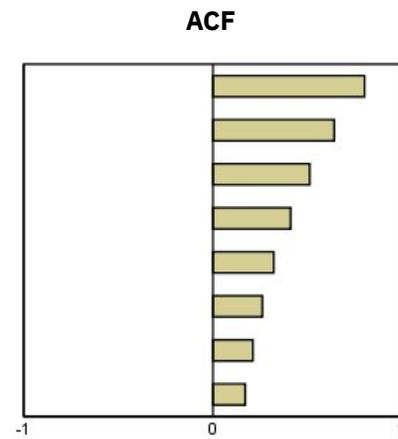


Tableau 7. ARIMA(0,1,0) (séries intégrées)



Remarques

Le présent document a été développé pour des produits et des services proposés aux Etats-Unis. et peut être mis à disposition par IBM dans d'autres langues. Vous pouvez toutefois devoir détenir une copie du produit ou une version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Contactez votre interlocuteur IBM pour plus d'informations sur les produits et services disponibles dans votre région. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service IBM puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits décrits dans le présent document. La remise de ce document ne vous accorde aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785U.S.A.

Les informations sur les licences concernant les produits IBM utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante :

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

19-21, Nihonbashi-Hakozakicho, Chuo-kuTokyo 103-8510, Japon

LE PRESENT DOCUMENT EST LIVRE "EN L'ETAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Il est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, apporter des améliorations et des modifications aux produits et aux logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci) et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785U.S.A.

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles pour ce dernier sont fournis par IBM d'après les termes du contrat client IBM, des conditions internationales d'utilisation de logiciels IBM ou de tout contrat équivalent entre nous.

Les données de performances et les exemples de clients ne sont présentés qu'à des fins d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins d'illustration ou d'explication uniquement. Toute ressemblance avec des noms de personnes et de sociétés serait purement fortuite.

LICENCE DE COPYRIGHT :

Le présent logiciel contient des exemples de programmes d'application en langage source destinés à illustrer les techniques de programmation sur différentes plateformes d'exploitation. Vous avez le droit de copier, de modifier et de distribuer ces exemples de programmes sous quelque forme que ce soit et sans paiement d'aucune redevance à IBM, à des fins de développement, d'utilisation, de vente ou de distribution de programmes d'application conformes aux interfaces de programmation des plateformes pour lesquelles ils ont été écrits ou aux interfaces de programmation IBM. Ces exemples n'ont pas été testés en détails, ni dans toutes les conditions. C'est pourquoi IBM ne peut pas garantir ou assurer la fiabilité, la serviceabilité ou le fonctionnement de ces programmes. Les modèles de programmes sont fournis "en l'état", sans garantie d'aucune sorte. IBM ne sera en aucun cas responsable des dommages liés à l'utilisation de ces programmes.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

© Copyright IBM Corp. 2021. Des parties de ce code sont proviennent d'IBM Corp. Programmes exemples.

© Copyright IBM Corp. 1989 - 2021. All rights reserved.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp., dans de nombreux pays. Les autres noms de services et de produits peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à www.ibm.com/legal/copytrade.shtml.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-Unis et/ou dans d'autres pays.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques d'Intel Corporation ou de ses filiales aux Etats-Unis et dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Java ainsi que tous les logos et toutes les marques incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

Index

A

- ACF
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
 - tracés des processus ARIMA réels [49](#)
- analyse de séries temporelles
 - modèles de causalité temporelle [24](#)
- Analyse harmonique [22](#)
- Appliquer des modèles de séries temporelles
 - enregistrement de modèles réestimés en XML [19](#)
 - enregistrement de prévisions [19](#)
 - fonction d'autocorrélation partielle [16](#)
 - Fonction d'autocorrélation partielle résiduelle [16](#), [17](#)
 - Fonction d'autocorrélation résiduelle [17](#)
 - intervalles de confiance [17](#), [20](#)
 - modèles les plus ou les moins ajustés [18](#)
 - nouveaux noms de variable [19](#)
 - paramètres de modèle [16](#)
 - période d'estimation [14](#)
 - période de prévision [14](#)
 - prévisions [16](#)
 - Prévisions [17](#)
 - réestimer les paramètres de modèle [14](#)
 - statistiques dans tous les modèles [16](#)
 - Statistiques dans tous les modèles [17](#)
 - statistiques de Ljung-Box [16](#)
 - statistiques de qualité d'ajustement [17](#)
 - statistiques de qualité d'ajustement [16](#)
 - valeur manquante [20](#)
 - valeurs d'ajustement [17](#)

B

- BIC (critère d'information bayésien) normalisé
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)

D

- Désaisonnalisation
 - Calcul de moyennes mobiles [20](#)
 - Création de variables [21](#)
 - enregistrement des nouvelles variables [21](#)
 - Hypothèses [20](#)
 - Modèles [20](#)
- données d'historique
 - dans Time Series Modeler [11](#)
- données historiques
 - dans Appliquer des modèles de séries chronologiques [17](#)

E

- enregistrer
 - modèles réestimés en XML [19](#)
 - nouveaux noms de variable [12](#), [19](#)
 - prévisions de modèle [12](#), [19](#)
 - spécifications de modèles au format XML [12](#)
- Erreur Absolue Maximum
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
- Erreur Absolue Moyenne
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
- erreur de pourcentage absolue maximum
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
- erreur de pourcentage absolue moyenne
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
- erreur quadratique moyenne
 - dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
 - dans Time Series Modeler [10](#), [11](#)
- événements
 - dans Time Series Modeler [6](#)

F

- fonction d'autocorrélation partielle
 - dans Appliquer des modèles de séries chronologiques [17](#)
 - dans Time Series Modeler [10](#)
 - tracés des processus ARIMA réels [49](#)
- fonction d'autocorrélation
 - dans Appliquer des modèles de séries chronologiques [16](#)
 - dans Time Series Modeler [11](#)
- fonction d'autocorrélation partielle
 - dans Appliquer des modèles de séries chronologiques [16](#)
 - dans Time Series Modeler [11](#)
- fonctions d'autocorrélation
 - dans Appliquer des modèles de séries chronologiques [17](#)
 - dans Time Series Modeler [10](#)
 - tracés des processus ARIMA réels [49](#)
- fonctions de transfert
 - ordres de dénominateur [8](#)
 - ordres de différence [8](#)
 - ordres de numérateur [8](#)
 - ordres saisonniers [8](#)
 - retard [8](#)

I

intervalles de confiance
dans Appliquer des modèles de séries chronologiques [17](#), [20](#)
dans Time Series Modeler [11](#), [13](#)

L

langage XML
enregistrement de modèles réestimés en XML [19](#)
enregistrement des modèles de séries chronologiques au format XML [12](#)

M

MAE
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

MAPE
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

MaxAE
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

MaxAPE
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

modèle de lissage exponentiel amorti [6](#)
modèle de lissage exponentiel de Brown [6](#)
modèle de lissage exponentiel de Holt [6](#)
modèle de lissage exponentiel de Winters
additif [6](#)
multiplicatif [6](#)

modèle de lissage exponentiel saisonnier simple [6](#)
modèle de lissage exponentiel simple [6](#)

Modèles
ARIMA [3](#)
lissage exponentiel [3](#), [6](#)
Modélisateur expert [3](#)

modèles ARIMA
fonctions de transfert [8](#)
valeurs éloignées [9](#)

modèles de causalité temporelle [24](#), [25](#), [27–31](#), [33–35](#)

modèles de lissage exponentiel [3](#), [6](#)

Modélisateur expert
limitation de l'espace de modèle [6](#)
valeurs éloignées [6](#)

N

Nom de variable
dans Time Series Modeler [12](#)

noms de modèle
dans Time Series Modeler [13](#)

noms de variable
dans Appliquer des modèles de séries chronologiques [19](#)

O

observations restantes [2](#)

P

PACF
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)
tracés des processus ARIMA réels [49](#)

paramètres de modèle
dans Appliquer des modèles de séries chronologiques [16](#)
dans Time Series Modeler [10](#)

période d'estimation
dans Appliquer des modèles de séries chronologiques [14](#)
dans Time Series Modeler [3](#)

période de prévision
dans Appliquer des modèles de séries chronologiques [14](#)
dans Time Series Modeler [3](#), [13](#)

période de validation [2](#)

période historique [2](#)

périodicité
dans Time Series Modeler [6](#), [8](#)

prévision de modèle de causalité temporelle [36–39](#), [41](#)

prévisions
dans Appliquer des modèles de séries chronologiques [16](#)
dans Time Series Modeler [11](#)

Prévisions
dans Appliquer des modèles de séries chronologiques [17](#)
dans Time Series Modeler [10](#)

Q

qualité d'ajustement
dans Appliquer des modèles de séries chronologiques [16](#)
dans Time Series Modeler [11](#)

Qualité de l'ajustement
dans Appliquer des modèles de séries chronologiques [17](#)
dans Time Series Modeler [10](#)
définitions [48](#)

R

R2
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

R2 stationnaire
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

réestimer les paramètres de modèle
dans Appliquer des modèles de séries chronologiques [14](#)

résiduels

résiduels (*suite*)
dans Appliquer des modèles de séries chronologiques [16](#)
dans Time Series Modeler [11](#)

Résidus
dans Appliquer des modèles de séries chronologiques [17](#)
dans Time Series Modeler [10](#)

RMSE
dans Appliquer des modèles de séries chronologiques [16](#), [17](#)
dans Time Series Modeler [10](#), [11](#)

S

scénarios de modèle de causalité temporelle [42–47](#)

Statistique de Ljung-Box
dans Time Series Modeler [10](#)

statistiques de Ljung-Box
dans Appliquer des modèles de séries chronologiques [16](#)

T

Time Series Modeler
ARIMA [3](#), [8](#)
enregistrement de prévisions [12](#)
enregistrement des spécifications de modèles au format XML [12](#)
événements [6](#)
fonction d'autocorrélation partielle [11](#)
Fonction d'autocorrélation partielle résiduelle [10](#), [11](#)
Fonction d'autocorrélation résiduelle [10](#)
fonctions de transfert [8](#)
intervalles de confiance [11](#), [13](#)
lissage exponentiel [3](#), [6](#)
modèles les plus ou les moins ajustés [12](#)
Modélisateur expert [3](#)
noms de modèle [13](#)
nouveaux noms de variable [12](#)
paramètres de modèle [10](#)
période d'estimation [3](#)
période de prévision [3](#), [13](#)
périodicité [6](#), [8](#)
prévisions [11](#)
Prévisions [10](#)
Statistique de Ljung-Box [10](#)
statistiques dans tous les modèles [11](#)
Statistiques dans tous les modèles [10](#)
statistiques de qualité d'ajustement [10](#)
statistiques de qualité d'ajustement [11](#)
transformation de séries [6](#), [8](#)
valeurs d'ajustement [11](#)
valeurs éloignées [6](#), [9](#)
valeurs manquantes [13](#)

Tracés spectraux
Analyse spectrale bivariée [22](#)
Centrage de transformation [22](#)
fenêtres spectrales [22](#)
Hypothèses [22](#)

transformation log
dans Time Series Modeler [6](#), [8](#)

transformation par log népérien

transformation par log népérien (*suite*)
dans Time Series Modeler [6](#), [8](#)

transformation racine carrée
dans Time Series Modeler [6](#), [8](#)

V

valeur extrême additive
dans Time Series Modeler [6](#), [9](#)

valeur extrême additive saisonnière
dans Time Series Modeler [6](#), [9](#)

valeur extrême corrective additive
dans Time Series Modeler [6](#), [9](#)

valeur extrême d'innovation
dans Time Series Modeler [6](#), [9](#)

valeur extrême de changement de niveau
dans Time Series Modeler [6](#), [9](#)

valeur extrême de tendance locale
dans Time Series Modeler [6](#), [9](#)

valeur extrême transitoire
dans Time Series Modeler [6](#), [9](#)

valeur manquante
dans Appliquer des modèles de séries chronologiques [20](#)

valeurs d'ajustement
dans Appliquer des modèles de séries chronologiques [17](#)
dans Time Series Modeler [11](#)

valeurs éloignées
modèles ARIMA [9](#)
Modélisateur expert [6](#)

valeurs extrêmes
définitions [49](#)

valeurs manquantes
dans Time Series Modeler [13](#)

