

A la Une du Support

RÉALISER DES PRÉVISIONS PERTINENTES GRÂCE AU MODULE SAS® HIGH-PERFORMANCE FORECASTING (PARTIE II)

Dans la précédente édition de cette newsletter, vous avez pu lire un article qui traitait du module SAS® High-Performance Forecasting et en particulier des deux principales procédures HPFDIAGNOSE et HPFENGINE. Dans cette suite, nous parlerons des autres procédures qui s'articulent autour des deux précédentes et qui nous permettent de personnaliser l'analyse de nos séries

Caractéristiques :

Catégories : SAS® High-Performance

Forecasting

OS : Windows, Unix, z/OS Version : SAS® 9.3 et 9.2 Vérifié en septembre 2012

temporelles en ajoutant des éléments (événements, réconciliation,...).

Table des matières

| | Procédures HPF: rappel du concept | 1 |
|----|---|---|
| Le | es procédures HPF auxiliaires | 3 |
| | PROC HPF | |
| | Les PROC HPFxxxSPEC | |
| | PROC HPFSELECT | |
| | PROC HPFEVENTS | |
| | PROC HPFRECONCILE | |
| | En cas de difficulté | 7 |
| | Eléments à transmettre au Support Clients | 7 |
| | Conclusion | 7 |
| | | |

Procédures HPF: rappel du concept

Ces procédures ont pour but de générer des prévisions de grande qualité pour un large éventail de modèles et de données.

Elles peuvent être utilisées :

- de manière automatique, via SAS® Forecast Studio.
- par un utilisateur averti, via du code.

Leur utilisation permette d'obtenir des prévisions de séries temporelles plus précises qu'avec SAS/ETS®. En effet, ces procédures offrent une large panoplie d'outils et de cas de figure supportés (événements, réconciliation,...).

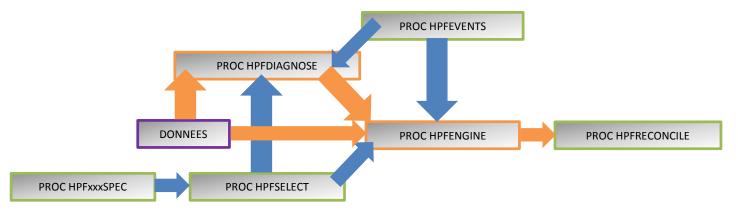
Chacune des procédures offre la possibilité de choisir parmi une panoplie d'options et de paramètres qui sont déterminés ou choisis automatiquement lors de l'utilisation de la solution « embarquée ». Il est ainsi possible de personnaliser de manière très pointue et précise les différentes prévisions que vous voulez effectuer.

Leur fonctionnement global est simple mais il est possible que vous soyez amené à utiliser pas mal de procédures et d'options si vous voulez une estimation pertinente de vos données. Des procédures permettent de créer des répertoires contenant des spécifications de modèles (par exemple, PROC HPFxxxSPEC), ensuite les meilleurs modèles de chaque catégorie sont conservés (lissage, ARIMAX, demande intermittente et modèle à composantes inobservées). On peut enfin faire un diagnostic sur la série, l'estimer et la prévoir, en incorporant des options liées aux événements et aux réconciliations.

Ces procédures, ainsi que la solution SAS Forecast Studio, ont été conçues dans le but de pouvoir prévoir un grand nombre de séries de manière industrielle et ainsi traiter plus efficacement les grandes quantités de données qui sont désormais disponibles. De ce fait, comme nous le détaillerons dans les deux sections suivantes, la mise en production (c'est-à-dire la création des différentes

spécifications de modèles) des toutes premières séries peut être assez longue et paraître fastidieuse. Cependant, cette tâche permettra de réaliser ensuite des prévisions rapidement et très efficacement. Un référentiel de modèles est créé et est alors disponible et utilisable pour toutes les autres analyses de données temporelles.

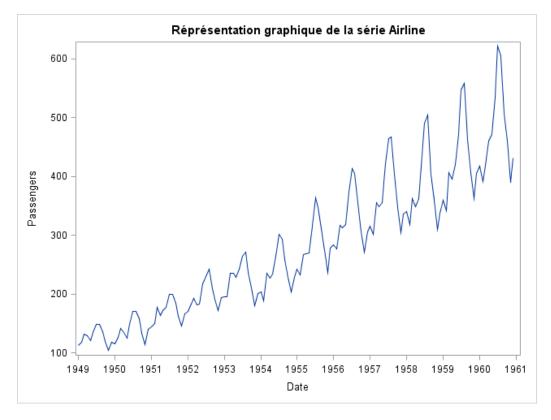
Ci-dessous une représentation globale des interactions entre les différentes procédures du module HPF.



NB: en orange, le parcours des données et, en bleu, celui des spécifications (modèle, événements,...)

Nous allons revenir plus en détail sur les procédures auxiliaires et leurs interactions dans les sections à venir.

Nous travaillons sur la série Airline qui traite de l'évolution du nombre de passagers (la même série que pour l'article précédent).



LES PROCÉDURES HPF AUXILIAIRES

Les deux procédures HPFDIAGNOSE et HPFENGINE que nous avons présentées dans l'article précédent sont les plus importantes (puisqu'elles permettent le diagnostic et les prévisions des différentes séries) mais il en existe d'autres, qui augmentent la capacité de personnalisation de ce type de modèle, sur les séries temporelles.

Dans cette partie nous allons détailler les autres procédures HPF.

PROC HPF

La procédure HPF (High-Performance Forecasting) fournit un moyen rapide et automatique pour générer des prévisions pour plusieurs séries temporelles ou données transactionnelles en une seule étape. Il est possible d'utiliser les principaux modèles de lissage (simple, double, linéaire, tendance amortie (Dramped), saisonnier ou Winters), les transformations courantes (log, racine carrée, logistique ou Box-Cox) et aussi des modèles de demande intermittente.

Les tables de sorties possibles sont des statistiques sommaires des séries, des prévisions et limites de confiance, des paramètres estimés et des statistiques d'ajustement.

La procédure HPF peut prévoir des données de séries temporelles, dont les observations sont espacées de manière identique dans le temps par mois, semaine, jour, ..., et aussi des données transactionnelles, dont les observations ne sont pas espacées par rapport à un intervalle de temps particulier. Pour les données transactionnelles, les données sont accumulées sur la base d'un intervalle de temps spécifique pour former une série temporelle. La procédure HPF peut aussi effectuer une analyse de tendance et de saisonnalité sur les données transactionnelles.

```
/*PROC HPF*/

□ Proc hpf data=tmpl.airline out=sortie outest=estimateur outfor=prevision outprocinfo=procnfo outstat=stat outsum=sum outseason=saison; forecast _ALL ; id Date interval=month; run;
```

Les options de sortie permettent de récupérer différentes tables :

- la table initiale à laquelle sont ajoutées les prévisions,
- la table initiale avec des informations sur les estimateurs,
- la table initiale avec les prévisions,
- une table avec les informations relatives au déroulement de la procédure, les statistiques d'ajustement,
- une table avec un résumé des prévisions,
- et une table avec les informations relatives à la saisonnalité.

Les PROC HPFxxxSPEC



Il existe plusieurs procédures de ce type dont les fonctionnements et usages sont proches (HPFARIMASPEC, HPFESMSPEC, HPFUCMSPEC, HPFIDMSPEC et HPFEXMSPEC), c'est la raison pour laquelle elles sont regroupées dans cette section. Nous allons en expliquer le principe pour une seule (HPFARIMASPEC, qui est peut-être la plus complexe de surcroît). Cette procédure travaille sur les modèles ARIMA alors que les quatre autres travaillent respectivement sur les lissages exponentiels, les modèles à composantes inobservées, les modèles à demande intermittente et les modèles externes.

La procédure HPFARIMASPEC est utilisée pour créer un fichier de spécifications/caractéristiques de modèle ARIMA (AutoRégressif Intégré Moyenne Mobile). La sortie de cette procédure est un fichier qui stocke la spécification du modèle destiné ARIMA. Ce fichier de spécification XML peut être utilisé

à des fins différentes, par exemple, pour être sauvegardé et utilisé ultérieurement par la procédure HPFENGINE. On peut spécifier des modèles ARIMA très généraux, tous les modèles qui peuvent être analysés avec la procédure ARIMA peuvent être spécifiés. Par ailleurs, la spécification du modèle peut inclure des transformations de séries telles que des transformations log ou Box-Cox.

Ainsi, une utilisation intéressante de cette procédure serait de créer tous les modèles vraisemblables qu'il est possible d'imaginer pour se créer une base de modèles complète. Cette base pourrait être utilisée pour prévoir les séries de nos différentes études futures. Il est conseillé de créer une base de modèles la plus grande et la plus exhaustive possible. Ainsi, ce travail ne sera à effectuer que lors de votre première étude et pourra être réutilisé pour les suivantes. Il est ainsi possible de donner un libellé et une description à vos spécifications de modèles. Un libellé du type « Modèle ARIMA(1,1,1) \times SARIMA(0,1,0) » parlant peut permettre une recherche et une application rapide.

Voici des exemples d'utilisation des procédures HPFARIMASPEC et HPFESMSPEC où les principales options communes sont reprises.

```
/*PROC HPFARIMASPEC*/

□ Proc hpfarimaspec name=arima_1 label="ARIMA(1,1,0)x SARIMA(0,1,1)" repository=work.mesmodeles; forecast symbol=passagers dif=(1,12) p=1 q=12; estimate method=ml maxiter=500; run;

/*PROC HPFESMSPEC*/

□ Proc hpfesmspec name=esm_1 label="Meilleur modèle de lissage "repository=work.mesmodeles; esm criterion=MAPE method=best transform=auto; run;
```

Avec ces procédures, il faut simplement spécifier le nom et le label/libellé de votre modèle (NAME= et LABEL=) et le répertoire où il sera stocké (REPOSITORY=).

NB: les instructions diffèrent d'un type de modèle à l'autre et ne sont pas détaillées dans cet article, la documentation en ligne explique très bien les différentes options possibles.

PROC HPFSELECT

PROC HPFDIAGNOSE

PROC HPFSELECT

PROC HPFENGINE

La procédure HPFSELECT permet de contrôler le processus de sélection de modèles de prévision en définissant des listes de modèles de prévision candidats. En utilisant les listes de sélection de modèles par la procédure HPFSELECT, on peut contrôler le modèle de prévision ou les modèles que le module SAS® High-Performance Forecasting utilise pour prévoir des séries temporelles particulières.

La procédure HPFSELECT crée des fichiers de sélection de modèles et les stocke dans un répertoire pour une utilisation ultérieure par d'autres procédures, la procédure HPFENGINE par exemple. Ces fichiers de sélection de modèles listent les spécifications de modèles précédemment créées et stockées dans un répertoire de modèles par les procédures HPFARIMASPEC, HPFESMSPEC, HPFEXMSPEC ou HPFUCMSPEC.

En utilisant la procédure HPFSELECT, d'autres options peuvent être spécifiées permettant de contrôler le processus de sélection de modèles de prévision. Cette dernière procédure, associée à celles de la section précédente, permet, une fois cette étape de création des spécifications passée, de diagnostiquer et de prévoir très rapidement n'importe quelle nouvelle série de données. C'est le principal avantage de ce module en ce qui concerne les séries temporelles sous SAS.

```
/*PROC HPFSELECT*/

Proc hpfselect repository=mesmodeles name=maselection label="Liste simple avec lissage et modèle ARIMA simple "; select criterion=mape holdout=12; specification arima_1 esm_1; run;

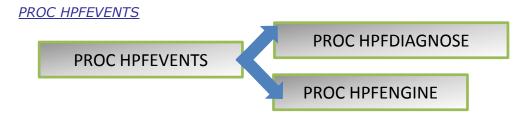
Proc hpfselect repository=mesmodeles name=maselection2; spec arima_1 esm_1; select criterion=mape choose=esm_1;/*choose=<> court-circuite la sélection en choisissant le modèle désigné*/ run;
```

On retrouve dans cette procédure les options REPOSITORY, NAME et LABEL qui servent respectivement à spécifier l'emplacement du répertoire qui contient les spécifications de modèles, le nom de la liste de modèles qui sera créée et le libellé associé.

L'instruction SELECT permet de déterminer les options qui seront utilisées lors de la sélection du meilleur modèle lorsque cette liste sera utilisée (un échantillon out-of-sample de 12 observations, HOLDOUT=12, avec le critère MAPE, CRITERION=mape, pour le premier exemple de code ci-dessus). Il est également possible de choisir le modèle qui sera utilisé. La seconde liste, MASELECTION2, utilise l'option CHOOSE= qui nous permet de spécifier le modèle à retenir, quelles que soient les données et les statistiques associées à ce modèle et aux autres (ainsi le modèle choisi ne sera pas forcément le meilleur mais celui que l'on souhaitait).

L'instruction SPECIFICATION (ou SPEC) permet de déterminer les modèles candidats qui seront contenus dans cette liste de modèles et ainsi étudiés lorsque la liste déterminera le meilleur modèle dans les procédures HPFDIAGNOSE ou HPFENGINE. Il s'agit ici des noms des spécifications de modèles qui ont été préalablement effectuées avec les procédures HPFxxxSPEC.

Bien d'autres options et instructions existent pour cette procédure (et pour les autres présentées dans cet article). En faire la liste et détailler le fonctionnement de manière exhaustive serait inutile. En effet, chacun cherchera les options nécessaires à son étude en fonction de ses données, son cahier des charges et des analyses qu'il souhaite mener.



La procédure HPFEVENTS fournit un moyen de créer et de gérer les événements associés à des séries temporelles pour des fins d'analyse. La procédure peut créer des événements, lire des événements à partir d'une table de données d'événements, écrire des événements dans une table de données d'événements, et créer des variables muettes basées sur ces événements si une information à jour est fournie.

Un événement SAS est utilisé pour modéliser tout incident qui perturbe le flux normal du processus qui génère les séries chronologiques. Des exemples d'événements communément utilisés incluent les catastrophes naturelles, les promotions de vente au détail, les grèves, les campagnes publicitaires, des changements de politiques et des erreurs d'enregistrement de données.

Un événement a un nom de référence, une ou des dates associées à l'événement et un ensemble de qualificatifs. L'événement existe indépendamment de toute série temporelle ; toutefois, l'événement peut être appliqué à une ou plusieurs séries temporelles. Lorsque l'événement est appliqué à une série temporelle, une variable muette (dichotomique) est générée et peut être utilisée pour analyser l'impact de l'événement sur les séries temporelles.

Il y a beaucoup d'avantages à utiliser PROC HPFEVENTS :

- Les variables muettes générées par PROC HPFEVENTS sont automatiquement prolongées, écourtées ou modifiées, et des observations sont ajoutées et/ou supprimées de la série.
 PROC HPFEVENTS peut être utilisée pour plusieurs séries temporelles ou pour différents « spans » de la même série.
- PROC HPFEVENTS peut être utilisée pour définir les variables muettes qui fonctionnent aussi bien pour les séries temporelles avec divers intervalles de temps, telles que les données hebdomadaires ou mensuelles. La même définition EVENT peut modéliser des données quotidiennes ou des totaux hebdomadaires.
- Les définitions d'événements peuvent être stockées dans une table. Les définitions d'événements peuvent être changées ultérieurement, de nouveaux éléments ajoutés ou des variables muettes supplémentaires générées à partir d'une table existante.
- Les définitions d'événements stockées dans une table peuvent être transmises directement à PROC HPFENGINE et PROC HPFDIAGNOSE.
- Les définitions d'événements SAS prédéfinis peuvent être consultées directement à partir de PROC HPFENGINE et PROC HPFDIAGNOSE.
- PROC HPFEVENTS peut générer une table qui peut être utilisée dans d'autres procédures telle que PROC REG. Comme les données sont ajoutées ou supprimées de la série chronologique, PROC HPFEVENTS peut générer automatiquement de nouvelles variables muettes comme requis.
- PROC HPFEVENTS reconnaît les dates et les variables prédéfinies. Ainsi, les événements qui impliquent des fêtes comme Pâques et Thanksgiving peuvent être intégrées dans la modélisation facilement, même si les dates des événements changent d'année en année.

La série sur laquelle se base notre exemple d'application ne contient pas de choc comme le montre le graphique représentant la série au début de cet article. Ainsi, notre exemple Airline n'est pas idéal dans ce contexte. De ce fait, pour que vous ayez un aperçu du type de code de cette procédure, voici un exemple de création d'événement que l'on trouve dans la documentation.

```
/*PROC HPFEVENTS*/

*Pas d'événement pour cette exemple;

*Exemple issu de la documentation;

□ proc hpfevents data=work_intv lead=12;

id date interval=month;

eventdef promotion= '01oct80'd / TYPE=LS;

eventdata out= evdsout1 (label='list of events');

run;
```

L'option DATA= permet de spécifier la table dans laquelle se trouve les variables utilisées dans les instructions VAR, ID et BY, tandis que l'option LEAD= permet de paramétrer l'horizon de l'extension des variables.

L'instruction ID et l'option INTERVAL= fonctionnent de manière similaire à ce qui se passe avec les autres procédures où elles se trouvent, comme la procédure HPF.

L'instruction EVENTDEF permet de spécifier la date (1^{er} octobre 1980), la variable sur laquelle l'événement se base (« promotion ») et le type d'effet (TYPE=).

La dernière instruction, EVENTDATA, permet à l'utilisateur de choisir la table dans laquelle l'événement sera exporté avec le libellé associé.

PROC HPFRECONCILE

PROC HPFENGINE

PROC HPFRECONCILE

Lorsque les données sont organisées de façon hiérarchique, il y a souvent des contraintes numériques qui lient les séries à différents niveaux de la hiérarchie. Par exemple, le total des ventes d'un produit par une entreprise de vente au détail est la somme des ventes de ce même produit dans tous les magasins qui appartiennent à l'entreprise.

Il semble naturel d'exiger que ces mêmes contraintes soient prises en compte lors du calcul des valeurs prédites. Imposer de telles contraintes au cours du processus de prévision peut être difficile voire impossible. Les prévisions sont réalisées à chaque niveau de manière indépendante, aussi ces contraintes ne sont pas respectées – de fait, il est nécessaire d'appliquer un processus de réconciliation.

La procédure HPFRECONCILE réconcilie les prévisions des données de séries temporelles à deux niveaux différents d'une hiérarchie.

La procédure HPFRECONCILE peut désagréger les prévisions du niveau supérieur des prévisions ou agréger les prévisions du niveau inférieur. La procédure permet à l'utilisateur de spécifier la direction et la méthode de la réconciliation, des contraintes d'égalité et des bornes sur les valeurs réconciliées. Pour désagréger une prévision à des niveaux inférieurs de la hiérarchie, la procédure utilise la proportion. Par exemple, si la prévision globale est de 200 et que les deux prévisions du niveau inférieur sont 30 et 70, les prévisions réconciliées sont respectivement 60 et 140.

Si l'on a choisi la méthode inverse, l'agrégation (ce qui signifie que l'on a plus confiance dans les prévisions des niveaux inférieurs par rapport à celle globale), il suffit d'accumuler les niveaux inférieurs pour connaître la valeur de notre prévision réconciliée. Dans notre exemple avec 30 et 70 au niveau inférieur, la prévision globale réconciliée serait de 100.

Notre série d'exemples ne contient pas de hiérarchie, nous ne pouvons donc pas expliciter le code de cette procédure. Cependant, dans le code joint, vous trouverez un exemple issu de la documentation sur une série de ventes par Région et par Produit permettant de comprendre le fonctionnement de cette procédure. Celui-ci est très simple, mais il est essentiel de comprendre que la réconciliation se fait niveau par niveau.

Si vous avez trois niveaux de hiérarchie, vous aurez deux réconciliations et donc deux appels de la procédure HPFRECONCILE.

En cas de difficulté

Eléments à transmettre au Support Clients

Si vous rencontrez des difficultés lors de l'utilisation de ce module, vous pouvez nous écrire à support@sas.com, en attachant à votre message l'erreur enregistrée dans votre journal.

Conclusion

Le module SAS® High-Performance Forecasting est LE module SAS pour faire des prévisions de qualité dans un mode industriel. Il supporte la plupart des méthodes de prévisions (ARIMA, ESM, IDM, UCM). On peut créer une spécification de modèle aisément, la stocker dans une liste et utiliser cette dernière pour toutes les analyses temporelles suivantes. De plus, on peut incorporer un événement (promotion, catastrophe naturelle,...) dans plusieurs séries de manière automatique. Enfin, on peut réconcilier des séries en quelques lignes de programme contre plusieurs dizaines avec SAS/ETS.

Dans <u>ce fichier compressé</u>, vous trouverez un exemple d'application ainsi que la table associée.

Jérémy NOEL Consultant Support Clients SAS France