# Soutien du cours

jeudi 6 décembre 2018

### PUBLICATIONS DES RECHERCHES ET DONNEES CONJONCTURELLES ET INTRODUCTION AU TRAITEMENT DES SERIES TEMPORELLES A L'AIDE L'ECONOMETRIE (module A10 (2))

### Contents

1.	Introduction
1.1.	Les documents du cours
1.2.	Contenu prévu
2.	Le KOF, comme exemple d'un centre de recherches conjoncturelles
2.1.	L'in stitut
2.2.	L'Organigramme
2.3.	Les Enquêtes conjoncturelles
2.4.	Les Enquêtes structurelles
2.5.	Enquête sur l'innovation
2.6.	KOF Panel des entreprises
2.7.	Prévisions & Indicateurs
2.8.	Un nouveau coup d'œil comparative sur les publications
3.	Le CNES et les sources disponibles conjoncturelles actuels et futures
3.1.	Les Données et les Projections du FMI du PIB Annuel 14
3.2.	Les Enquêtes Trimestrielles
3.3.	Les opinions de L'Enquête de Conjoncture trimestrielles 15
3.3.	Un exemple de données sectorielles: L'Enquête de Conjoncture : L'industrie 16
3.4.	Comment arriver à des prévisions ?
3.5.	Exercices préparés :
4.	Les Travaux en Groupe
4.1.	Les Schéma de Modélisation des Secteurs : Boites à outils conjoncturels
4.2.	Constitution des groupes
4.3.	Les travaux et Exercices du cours en groupe
4.4.	Préparations et Présentations des travaux en groupes :
5.	La Régression PCR et PLS avec XLSTAT
5.1.	Description
5.2.	Boîte de dialogue

5.3.	Résultats	
5.4.	Exemple du tutoriel: Régression PLS (moindres carrées partiels)	
5.5.	Exercices préparés	
5.6.	Les Exercices du cours en groupe	
Bibliogra	phie :	

# 1. Introduction

Ce document commence en encore une fois avec une description d'un exemple d'un centre de recherches conjoncturelles. Et pour varier de prendre les exemples de la France en a choisi le KOF de Zurich en Suisse avec l'avantage que leurs publications sont en français et peuvent être abonnées facilement de la Algérie et deuxièmement pour la qualité de travail et de la réputation de ce centre.

En va essayer, a l'exemple de KOF de continuer les exercices sur les informations sectorielles: énergie, agriculture, santé, social etc.. Dans la mesure de la comparabilité de la situation économique et politique.

La continuation de ce soutien du cours est consacrée à une préparation des traitements statistiques, sans aborder l'utilisation des logiciels prévus dans ce cours comme XLSTAT et EViews, la présentation de l'économétrie et à sa liaison avec la théorie économique. Nous abordons tout d'abord la notion de modèle ainsi que les différentes étapes de la modélisation. L'apport de l'économétrie en tant qu'outil de validation est étudié plus tard.

Apres la présentation des méthodes du KOF et leur comparaison avec disponibilités des données en Algérie en reprend rapidement les techniques d'analyse et prévisions traités déjà dans le module A7.

Ce document reprend plusieurs champs de l'économétrie: régression simple, des hypothèses, modèle à décalage, analyse des séries temporelles, économétrie des variables qualitatives etc. Par ailleurs nous introduisons des méthodes standard de traitement des séries temporelles, les méthodes de désaisonnalisation et de lissage

Apres suit une invitation au travail en groupe et un schéma proposé de travail en group que nous estimons d'être utile pour une préparation de publications actuelles du CNES

Nous avons voulu, par une alternance systématique de cours et d'exercices, répondre à un besoin pédagogique qui est de mettre rapidement en pratique les connaissances théoriques et ainsi, d'utiliser de manière opérationnelle les acquis du cours ; les exercices sont inclus fréquemment. De surcroît, le recours à des logiciels, lors de la résolution des exercices, permet une découverte de ces outils et donne une dimension pratique.

Afin que les participants puissent eux-mêmes refaire les exercices, les données utilisées (sous format Excel) sont disponibles. Les logiciels XLSTAT sera utilisé toujours comme également les fonctions d'EXCEL.

1.1. Les documents du cours

Les documents EXCEL avec l'utilisation de XLSTAT

1.Fichier-..J\_ONS\_PIB ALGERIE\_KR...xlsx

2.Algérie\_BM\_World\_Development\_Indicators180212.xlsx

3.Algérie\_et\_Int\_GDP\_Data\_Extract\_From\_World\_Development\_Indicators18 0212.xlsx

4. AlgerieEtMonde\_UtilisationPétrole\_Data\_Extract\_From\_World\_Developmen t\_Indicators180212.xlsx

5. Fichierd'exercice du tutoriel: demoPLS.xls

Les documents du soutien du cours

- A. Un Bulletin d'exemple du KOF (KOF\_bulletin\_2017\_05\_fr.pdf)
- *B.* Le rapport de la Conjoncture du CNES actuel (Conjoncture-rapport-global CNES S12015.pdf)
- C Enquete\_Opinion\_Industrie\_3T\_17.pdf
- D Enquete\_Opinion\_Commerce\_3t\_16.pdf
- E. Introduction aux méthodes avances de la combinaison de la Régression et L'Analyse en Composantes Principales (logiciel XLSTAT Régression PLS180210.pdf)
- F Soutien du Cours de la Formation A7 (Formation SPRING\_A7\_Manuel Statistique171029.pptx)
- G. Manuel Statistique partiellement traduit: (HandbookStats\_partiel\_fr\_KR171029.docx et Formation SPRING\_A7\_Manuel Statistique171020.pptx)
- H. Introduction au logiciel XLSTAT (Formation SPRING\_A7\_XLSTAT\_171109.pptx)

# 1.2. Contenu prévu

Jour1: Récapitulation des expériences des participants de la formation. Récapitulation de la Modélisation macroéconomique et utilisation de XLSTAT. Présentation a l'exemple de KOF de continuer les exercices sur les informations sectorielles: énergie, agriculture, santé, social etc

- Jour 2: Travaux en groupes: Modélisation macroéconomique avec application des méthodes connues sur PIB et comparaison international. Présentation des Travaux en groupes
- Jour 3: Modélisation macroéconomique avec application des méthodes dans les secteurs du Rapport de Conjoncture Economique & Sociale du CNES Introduction de la théorie statistique avec l'utilisation de XLSTAT : La Régression PCR, PLS1 et PLS 2, PLS (PLS-DA)
- Jour 4: Travaux en groupes: Modélisation macroéconomique avec application des méthodes connues sur les secteurs du Rapport de Conjoncture Economique & Sociale du CNES, préférablement dans le domaine de travaux des participants.
- Jour 5: Continuation des Travaux en groupes: Modélisation macroéconomique avec application des méthodes connues.

Présentation des Travaux en groupes: Récapitulation des expériences des participants de la formation et prévision sur l'application de la modélisation pour le (s) Rapport (s) future(s) de Conjoncture.

# 2. Le KOF, comme exemple d'un centre de recherches conjoncturelles

### 2.1. L'institut

Le KOF (https://www.kof.ethz.ch/fr/), Centre de recherches conjoncturelles, fournit des informations tangibles dans le domaine de la recherche économique et conjoncturelle. Il établit une multitude de prévisions et d'indicateurs utiles à l'observation de la conjoncture. Sa base de données, issue d'enquêtes menées auprès des entreprises, est unique en Suisse. Ces données permettent au KOF de générer divers indicateurs (baromètre, indicateur de l'emploi, indicateur bancaire), qui reflètent le climat de l'économie suisse. Les résultats de ces enquêtes sont évalués chaque trimestre et communiqués au public, tout comme les prévisions conjoncturelles. Le KOF analyse également l'activité d'innovation des entreprises suisses, effectue des études sur les dépenses de santé et commente les derniers développements de l'économie. Il participe aussi (en collaboration avec le centre IFO de Munich) à l'élaboration des prévisions conjoncturelles allemandes, lesquelles servent de base au gouvernement allemand en ce qui concerne l'évolution future de l'économie.



### 2.2. L'Organigramme

Malheureusement l'organigramme n'existe qu'en anglais : En voit une liaison étroite avec le centre de recherche de l'ETH Zurich et leurs professorats de recherche lies à la conjoncture.

- 2.3. Les Enquêtes conjoncturelles
- Enquête conjoncturelle de l'industrie

L'enquête conjoncturelle menée par le KOF dans l'industrie consiste en un sondage mensuel effectué auprès de sociétés industrielles suisses volontaires provenant d'un grand nombre de branches industrielles. Chaque trimestre, des questions complémentaires sont posées. Les questionnaires portent principalement sur des estimations (questions qualitatives) relatives aux activités récentes, présentes et futures de l'entreprise.

**Toutes les entreprises participantes reçoivent chaque mois des résultats personnalisés,** permettant une comparaison avec d'autres branches. Ce rapport offre une information compacte et synoptique des résultats de la dernière enquête. De plus, les participants reçoivent chaque trimestre un rapport détaillé et commenté au sujet des résultats. Les membres de la Société suisse de recherche conjoncturelle (SGK) bénéficient également de ce service.

Le questionnaire peut être rempli en ligne. Les participants reçoivent le code utilisateur et le mot de passe personnel dès l'inscription. En peut aussi tester l'enquête en ligne sans engagement personnel. En peut également y participer en utilisant la version papier. En peut sélectionner le mode de prédilection au moment de l'inscription. Les modifications sont possibles à tout moment.

### • Enquête conjoncturelle de la construction

Les enquêtes conjoncturelles menées par le KOF dans le secteur du bâtiment consistent en un sondage mensuel effectué auprès d'entreprises du bâtiment suisses volontaires. Les questionnaires portent principalement sur des estimations (questions qualitatives) relatives aux activités récentes, présentes et futures de l'entreprise.

Chaque mois, les participants reçoivent par e-mail un bulletin d'information contenant les derniers résultats. Et chaque trimestre, ils reçoivent en outre un rapport détaillé et commenté, adressé exclusivement aux entreprises participant à l'enquête ainsi qu'aux membres de la SGK.

Le questionnaire peut être rempli en ligne etc.

### • Enquête conjoncturelle dans le commerce en gros

L'enquête conjoncturelle menée par le KOF dans le commerce de gros et soutenue par Commerce Suisse, consiste en un sondage trimestriel effectué auprès d'entreprises volontaires issues d'un grand nombre de branches de commerce de gros.

Le rapport trimestriel sur le commerce de gros contient les résultats détaillés et commentés de l'enquête. Il est adressé exclusivement aux entreprises participant à l'enquête..

Le questionnaire peut être rempli en ligne etc

### • L'enquête conjoncturelle menée auprès du commerce de détail

L'enquête conjoncturelle menée auprès du commerce de détail s'agit d'un sondage mensuel effectué auprès d'entreprises de divers secteurs du commerce de détail en Suisse. Les questionnaires portent principalement sur des estimations (questions qualitatives) relatives aux activités récentes, présentes et futures de l'entreprise.

Le rapport mensuel sur le commerce de gros contient les résultats détaillés et commentés de l'enquête.

Le questionnaire peut être rempli en ligne etc.

### • Enquête conjoncturelle de la restauration et de l'hébergement

L'enquête conjoncturelle menée par le KOF auprès de l'hôtellerie est un sondage trimestriel effectué auprès d'entreprises suisses volontaires du secteur hôtelier. L'enquête fait la distinction entre les domaines suivants :

- Restauration
- Hébergement

Le rapport trimestriel sur l'hôtellerie contient les résultats détaillés et commentés de l'enquête. Le questionnaire peut être rempli en ligne etc.

Les deux enquêtes suivantes en laisse a l'étude du lecteur :

- Bureaux d'études
- Services financiers et assurantiels
- Enquête conjoncturelle des sociétés de services

L'enquête conjoncturelle menée par le KOF auprès des sociétés de services est un sondage trimestriel effectué auprès d'entreprises suisses volontaires du secteur tertiaire, appartenant aux domaines suivants :

- Transport et stockage
- Information et communication
- Immobilier
- Professions libérales, services scientifiques et techniques
- Autres services économiques
- Santé et affaires sociales
- Art, détente et loisirs

Le rapport trimestriel sur le secteur des services contient les résultats détaillés et commentés de l'enquête.

Le questionnaire peut être rempli en ligne etc.

#### • Enquête sur les investissements

Cette enquête diffère fortement des autres mais peut-être importent comme maquette pour l'Algérie

L'enquête du KOF sur les investissements s'inscrit dans le domaine des enquêtes conjoncturelles du KOF EPF Zurich ; il se conçoit donc comme une enquête destinée à recenser en temps utile l'activité d'investissement des entreprises suisses.

Les investissements productifs sont soumis à des variations conjoncturelles considérables et fortement corrélés aux taux de croissance du PIB. L'enquête sur les investissements est effectuée

chaque année en automne et au printemps, dans le but de recenser en temps opportun l'évolution des investissements ; elle contient des questions sur les projets et les objectifs d'investissement, sur les facteurs d'influence ainsi que sur les motivations des investissements directs à l'étranger (cf. questionnaire). Le montant des investissements est demandé aussi bien en ce qui concerne l'activité passée que future. Il est ainsi possible d'en tirer des conclusions concernant la réalisation des projets d'investissement. Les entreprises sont interrogées sur les investissements productifs (équipement et construction) ainsi que sur les investissements dans la recherche et le développement. Cela permet une détermination profonde et précise du cycle actuel d'investissement.

Les participants à l'enquête proviennent d'un échantillonnage représentatif des entreprises privées de l'industrie, du bâtiment et des services. Environ 8000 entreprises sont inscrites à chaque enquête sur les investissements du KOF, et plus de 3000 d'entre elles en moyenne répondent au questionnaire. L'échantillon correspond en grande partie à celui de l'enquête sur l'innovation, afin de répondre à l'objectif de cohérence entre les participants aux enquêtes.

Afin de prendre en considération de l'évolution constante de la composition de l'économie suisse et de compenser le désistement d'entreprises participantes, chaque vague d'enquêtes intègre de nouvelles entreprises supplémentaires dans le panel d'entreprises.

# 2.4. Les Enquêtes structurelles

Les enquêtes structurelles comme le panel d'entreprises du KOF permettent d'obtenir des données sur les activités d'innovation à intervalles réguliers (tous les trois ans). Par ailleurs, des données sont également collectées à un rythme irrégulier (environ une fois par an) au sujet de divers aspects liés à l'évolution structurelle de l'économie suisse, notamment l'emploi des technologies d'information et de communication, l'évolution organisationnelle des entreprises, les qualifications de la main-d'œuvre, l'internationalisation de l'économie suisse et la création d'entreprise (thèmes traités jusqu'à présent).

# 2.5. Enquête sur l'innovation

Depuis 1990, à raison d'une fois tous les trois ans, le KOF effectue une enquête sur l'innovation pour le compte du Secrétariat d'Etat à l'économie (SECO). Cette enquête a d'abord pour objectif de faire le point de la capacité d'innovation de l'économie suisse (analyse descriptive) ; ensuite, elle permet d'en tirer des conclusions utiles à la politique économique (analyse nominative) ; enfin, les données collectées apportent un élément d'explication scientifique sur le processus d'innovation (analyse explicative).

Les données dont collectées dans le cadre d'une vaste enquête, menée auprès du panel d'entreprises du KOF. Les résultats des analyses fondées sur les données issues de cette enquête font l'objet d'un rapport de synthèse. Ce rapport est publié sur le site du SECO et en tant qu'étude du KOF. Le résumé du rapport est disponible électroniquement

# 2.6. KOF Panel des entreprises

Depuis 1996, le KOF dispose d'un panel d'entreprises qui repose sur un échantillon d'environ 6500 sociétés provenant du Registre des entreprises et des établissements (REE) de l'Office fédéral de la statistique (OFS). L'échantillon, qui couvre l'industrie, le bâtiment et la partie

commerciale du secteur des services, est structuré par secteur d'activité et catégorie de taille d'entreprise ; il est actualisé à intervalles réguliers.

La base de données d'adresses correspondante, qui contient non seulement de nombreuses caractéristiques structurelles mais aussi toutes les données nécessaires au bon déroulement des enquêtes (services spécifiques, contrôle des réponses, relances et sondage des entreprises ne répondant pas), a été régulièrement améliorée et actualisée au cours du projet.

Ce panel d'entreprises du KOF permet de mener des enquêtes régulières sur les activités d'innovation (tous les trois ans) et les projets d'investissement (chaque année) des entreprises suisses. Par ailleurs, des données sont collectées à intervalles réguliers (environ une fois par an) sur divers aspects de l'évolution structurelle de l'économie suisse, notamment l'emploi des technologies de l'information et de la communication, l'évolution organisationnelle des entreprises, les qualifications des salariés, l'internationalisation de l'économie suisse ainsi que la création d'entreprises.

### • Base de données

Les données sont collectées dans le cadre d'une vaste enquête auprès des entreprises qui constituent le panel d'entreprises du KOF (EPF Zurich).

### • Questionnaire

Le questionnaire de chaque enquête sur l'innovation consiste en une partie standard, qui s'inspire, pour les questions essentielles, des enquêtes similaires menées dans les pays de l'UE, et en une partie spécifique, consacrée à un thème d'actualité.

### • Résultats

Les résultats des analyses effectuées sur la base des enquêtes sur l'innovation sont publiés sous forme d'un rapport final (en langue allemande). Ce rapport peut être consulté sur le site du SECO et dans les études du KOF. Le résumé du rapport est disponible sous forme électronique en allemand et en français. De plus, le site du SECO fournit une sélection d'indicateurs de l'enquête sur l'innovation, qui représentent l'évolution de ces variables dans la durée, ventilée selon divers critères (taille de l'entreprise, région et catégorie économique).

### 2.7. Prévisions & Indicateurs

Le KOF publie régulièrement divers indicateurs et prévisions. Ceux-ci se basent en partie sur les résultats d'enquêtes conjoncturelles du KOF auprès des entreprises suisses, comme sur des données extérieures.

### Les Prévisions

### **KOF Prévisions conjoncturelles**

Chaque trimestre, le KOF présente ses prévisions concernant l'évolution de l'économie suisse. Il élabore une analyse approfondie du cadre économique et politique mondial, sur lequel se fondent des prévisions économiques détaillées. Ces prévisions modélisées se distinguent par un degré

élevé de désagrégation et fournissent des informations sur l'évolution actuelle et prévisible des principaux aspects de l'économie nationale. La publication écrite s'inscrit dans le cadre des analyses du KOF.

### KOF Prévisions pour le tourisme

Mandaté par le SECO, le KOF produit des prévisions biannuelles concernant le tourisme suisse. Les prévisions concernant le tourisme du KOF sont respectivement publiées au printemps et en automne.

#### **KOF Consensus Forecast**

Depuis septembre 2007, le KOF établit son traditionnel « Consensus Forecast ». A cet effet, il interroge des économistes actifs du secteur privé et de l'administration publique au sujet de leurs prévisions concernant l'évolution des principales données de l'économie suisse. Les résultats trimestriels sont publiés sous forme de prévisions consensuelles.

#### **KOF International Forecasts**

Les « International Forecasts » du KOF contiennent les prévisions trimestrielles relatives aux Comptes nationaux des espaces économiques de l'OCDE et de l'UE ainsi que de l'Allemagne, de la France, de l'Italie, de la Grande-Bretagne, du Japon, des Etats-Unis et de la Suisse.

#### KOF Prévisions des dépenses de santé

Le KOF établit deux fois par an des prévisions sur l'évolution des dépenses du système de santé publique suisse ; il examine les trois dimensions « prestations », « prestataires » et « financement » sur la base d'un modèle économétrique.

#### Les Indicateurs

- KOF Baromètre conjoncturel
- KOF Indicateur de l'emploi
- KOF Monetary Policy Communicator
- KOF Indicateur de la situation des affaires
- KOF Indice de la mondialisation
- KOF Surprise Indicator
- KOF Indicateur Construction
- KOF Youth Labor Market Index
- KOF Indicateurs de l'insécurité

Des indicateurs peuvent être intéressant pour l'Algérie mais le « KOF Indice de la mondialisation « est immédiatement d'intérêt : Comment se positionne l'Algérie dans le contexte international et en comparaison avec ses voisins :

L'indice de la mondialisation du KOF mesure les dimensions économiques, sociales et politiques de la mondialisation. Il permet d'observer l'évolution de la

2018 KOF	Globalisation Index	
Classeme	nts pour l'année 2015	
	Pays	Indice de globalisation du classement
1	Netherlands	90.24
2	Switzerland	89.70
3	Sweden	88.05
64	Armenia	67.09
65	Morocco	67.01
66	Hong Kong, China	65.89
71	Mongolia	65.12
72	Tunisia	64.86
73	Trinidad and Tobago	64.66
118	Gabon	53.19
119	Algeria	53.18
120	Rwanda	53.17

Klaus Röder 06/12/2018

mondialisation dans une série de pays sur une longue période. L'indice de la mondialisation du KOF 2016 est disponible pour 187 pays sur la période de 1970 à 2013, selon un calcul tenant compte de 23 variables. L'indice de la mondialisation du KOF comprend une composante économique, sociale et politique. Il mesure la mondialisation sur un barème de 1 à 100. Les valeurs des variables de base sont subdivisées en centiles. Ceci permet de lisser les extrêmes et de réduire les fluctuations sur la durée. Les données utilisées ont été actualisées pour les dernières années à partir des sources d'origine. Les nouveaux chiffres ne sont pas comparables à l'indice du KOF publié il y a un an, étant donné que la base de données a été mise à jour et recalculée également pour toutes les années précédentes. Par conséquent, les comparaisons avec les années précédentes évoquées dans le texte reposent sur la nouvelle méthode de calcul

Parmi la vaste gamme des publications de KOF, le Bulletin est peut-être le plus important, certainement le plus régulier. Regardons le contenu de cet exemplaire (en haut) :

Regardez quelques chapitres représentatifs :

### KOF Enquêtes conjoncturelles avec

G 1 : KOF Indicateur de la situation des affaires

et T 1 : Indicateur KOF de la situation des affaires en Suisse (soldes, valeurs désaisonnalisées)

	Avr. 16	Mai 16	Juin 16	Juil. 16	Août 16	Sep. 16	Oct. 16	Nov. 16
Secteur privé (total)	7.5	9.8	8.1	9.8	10.7	11.1	11.5	9.7
Industrie de transformation	-7.6	-6.2	-3.6	-8.6	-7.8	-7.7	-6.0	-9.8
Bâtiment	24.9	22.3	22.9	23.9	27.6	23.3	25.0	26.4
Bureaux d'études	45.2	46.7	43.9	45.5	46.2	46.1	45.7	42.1
Commerce de détail	-9.4	-10.6	-8.6	-11.8	-10.7	-8.5	-7.5	-11.7
Commerce de gros	-6.8	-	-	3.7	-	-	3.5	-
Services financiers	18.1	26.3	14.5	18.2	22.8	22.9	24.1	21.5
Hôtellerie	-16.8	-	-	-21.9	-	-	-17.1	-
Autres services	20.0			23.7			21.9	

et

### L'Évolution géographique

Et autres

Un thème à poursuivre dans le prochain module A12 (MAPINFO)

### 2.8. Un nouveau coup d'œil comparative sur les publications

Klaus Röder 06/12/2018



-17

-52

.89

-98

-76



Le KOF Bulletin est une publication mensuelle. Il vous tient informé des nouvelles évolutions conjoncturelles importantes, notamment des résultats d'enquêtes du KOF, présentées dans un contexte global. Le bulletin vous informe également des derniers résultats de la recherche du KOF. Le KOF Bulletin est gratuit et il parait en allemand, français et anglais. Exemple du KOF : Le bulletin du 2/2018- Contenu :

- 1. La diffusion de la numérisation
- 2. L'activité d'investissement en 2018
- 3. L'endettement dans la zone euro

### INDICATEURS DU KOF

4. KOF Indice de la mondialisation : diminution à l'échelle mondiale en 2015



Evidement c'est les premiers paragraphes de l'importance internationale et de l'importance des décisions avec les effets sur le niveau national et international. Plus important peut-être pour le CNES sont les





G 1 : Envisagez-vous d'accroître, de réduire ou de maintenir vos investissements l'an prochain par rapport à l'année en cours ? (Solde)



#### À propos de l'enquête

La demière enquête du KOF sur les investissements se fondait sur un panel de plus de 8000 entreprises, reproduisant la structure de l'économie suisse. Elle ne prend pas en considération le secteur public, les entreprises semi-publiques ainsi que les ménages. Au total, plus de 3000 entreprises ont répondu.

#### conjoncturel.

L'exemple du baromètre conjoncturel et une proposition pour l'Algérie et le CNES : Comme le KOF compare la conjoncture réelle (le PIB) avec les prévisions de la conjoncture. Cette méthode peut être utilisée pour l'Algérie. Naturellement il n'existe pas un Baromètre mensuel mais les prévisions

de l'ONS peuvent être utilisées au niveau trimestriel.

Finalement nous voyons un Tableaux des prévisions actuel du KOF :

							١	/ariatio	ns en	% par	rappor	t				
			au	trimest	re pré	cédent	(désai	sonna	lisé, sı	ur base	annue	elle)		à l'an	née préc	édente
	2008-		20	17			20	18			20	19		2017	2018	2019
	2016	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4			
Consommation privée	1.6	0.5	0.9	1.3	1.4	1.6	1.7	1.6	1.5	1.4	1.4	1.4	1.3	1.3	1.5	1.5
Consommation publique	1.8	0.8	1.1	-1.0	1.6	2.4	2.7	1.6	0.1	-0.1	0.5	0.8	1.0	1.2	1.6	0.6
Formation brute de capital fixe	1.6	3.9	2.1	-2.4	1.4	6.6	2.9	0.1	2.9	2.8	2.7	2.6	2.4	1.5	2.4	2.4
- Construction	2.3	1.0	3.2	1.0	1.2	2.4	2.0	2.2	1.8	1.6	1.4	1.0	0.6	1.6	1.9	1.6
<ul> <li>Machines et équipements</li> </ul>	1.2	5.8	1.4	-4.5	1.5	9.4	3.5	-1.2	3.5	3.5	3.5	3.5	3.5	1.4	2.8	2.9
Exportations (1)	2.5	0.5	1.0	8.6	4.8	4.6	4.6	4.4	3.1	3.0	3.2	3.6	3.8	3.0	4.8	3.5
– Biens (1), (2)	1.9	16.8	6.4	6.0	5.4	5.2	4.8	4.8	4.7	4.5	4.4	4.4	4.5	5.6	5.2	4.6
- Services	2.2	-32.2	-1.3	3.0	6.8	5.0	5.7	4.2	0.8	0.7	1.5	2.4	3.1	-0.9	4.4	2.0
Importations (1)	2.8	- <mark>2.6</mark>	17.3	7.0	4.7	3.6	4.5	2.8	2.9	3.5	3.6	3.7	3.5	4.0	5.1	3.4
- Biens (1)	1.4	-3.0	23.7	10.7	5.9	4.2	5.8	2.9	3.1	3.8	3.9	3.9	3.5	4.9	6.4	3.7
- Services	6.1	-2.1	6.8	0.5	2.5	2.5	2.1	2.7	2.5	3.0	3.2	3.3	3.5	2.5	2.5	2.9
Variation des stocks (3)	-0.2	2.0	6.4	0.5	0.5	-0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3	0.0
Produit intérieur brut (PIB)	14	03	11	2.0	23	21	25	22	20	1.8	16	1.8	2.0	0.8	22	10

# 3. Le CNES et les sources disponibles conjoncturelles actuels et futures

Nous disposons des plusieurs publications nationales, p.e. d'ONS (F5\_PIB ALGERIE\_KR...xlsx/PIB\_annuel)

.. et les publications et les prévisions internationales



### 3.1. Les Données et les Projections du FMI du PIB Annuel :

Source: FMI / IMF World Economic Outlook 2018 / ONS 2017

# 3.2. Les Enquêtes Trimestrielles

Algérie dispose des informations trimestrielles de changement du PIB ((F5\_PIB ALGERIE\_KR...xlsx/PIB\_annuel/PIB prix courant)) Les séries : PIB trimestriel :



#### Source: ONS 2017

Algérie dispose également des informations trimestrielles de l'opinion du changement:

# 3.3. Les opinions de L'Enquête de Conjoncture trimestrielles

### Les Principales Caractéristiques de l'Enquête de Conjoncture

Les enquêtes de conjoncture (enquêtes d'opinion) qui sont des enquêtes légères, rapides et qui couvrent aussi les domaines non traités par les enquêtes traditionnelles, sont conçues dans le but :

- d'une évaluation qualitative des paramètres de la croissance ;
- d'une connaissance rapide des tendances de l'activité ;
- de l'analyse et de la prévision conjoncturelle.

Le champ de l'enquête trimestrielle auprès des entreprises industrielles est de 514 entreprises et filiales dont : 250 entreprises publiques

- 264 entreprises privées

Le questionnaire traite les différents domaines liés à l'activité de l'entreprise (le niveau de l'activité, la demande et la distribution, la main d'œuvre, les équipements et la trésorerie). Les résultats sont disponibles 45 jours après la fin du trimestre avec des **taux de réponse variant de 30 à 45%**. Les réponses aux questions posées dans le cadre des enquêtes de conjoncture (enquêtes d'opinion) sont à deux modalités (oui - non) ou à trois modalités (hausse - stable - baisse) selon le type de questions posées. Elles sont codées sous forme de notations chiffrées et pondérées en fonction de l'importance de chaque entreprise au sein de sa branche d'activité (chiffre d'affaires de l'année précédente) et par les poids respectifs des branches (valeur ajoutée de l'année de base (1989) pour le calcul des valeurs synthétiques moyennes aux divers niveaux d'agrégation.

Les résultats sont tirés sous forme de tableaux statistiques dont les chiffres sont présentés en termes relatifs (%) dégageant ainsi des soldes d'opinion pour les réponses à trois modalités et des proportions pour les réponses à deux modalités.

Exemple : A la question suivante « Quelle est la tendance de votre production au cours des trois derniers mois ? ». Trois réponses possibles sont prévues « en hausse (+) », « stable (0) », « en baisse (-)». *Le solde d'opinion est calculé comme la différence* (?) entre la proportion d'entreprises qui ont répondu que leur tendance est en hausse et celles pour lesquelles la tendance est en baisse (sans tenir compte de la proportion d'entreprises qui ont répondu que leur tendance est stable)

Ex calcul:

 $16 = \ll + \gg$ ,  $30 = \ll 0 \gg$ ,  $48 = \ll - \gg -> 16 / 48 = 0.33 - 1 = -66.67\%$  (entre -100% et +100%) Désavantage : Le nombre de non-réponses n'a aucune influence sur les résultats Ex alternative I:

+	0	I	Non- réponses	Somme	Questionnaires	Résultat	Résultat	Formule
2	10	3	3	15	18	-5.56%	-5.56% sur 18	=(A2-C2)/((E2+D2))
2	10	3	18	15	33	-3.03%	-3.03% sur 33	=(A3-C3)/((E3+D3))
7	12	3	18	22	40	10.00%	10.00% sur 40	=(A4-C4)/((E4+D4))
8	12	8	18	28	46	0.00%	0.00% sur 46	=(A5-C5)/((E5+D5))
8	12	8	55	28	80	0.00%	0.00% sur 80	=(A6-C6)/((E6+D6))
		т	1 1	,	<u> </u>	1 /	1	

Avantages : Le nombre de non-réponses sont reflétés dans les résultats

Source: ONS et Klaus Röder 2017

3.3. Un exemple de données sectorielles: L'Enquête de Conjoncture : L'industrie Voyez document (C Enquete\_Opinion\_Industrie\_3T\_17.pdf)

### L Activité d'industrie privé et publique



Source: (C) Enquete\_Opinion\_Industrie\_3T\_2017.pdf

#### La Demande



Source: (C) Enquete\_Opinion\_Industrie\_3T\_2017.pdf

#### La Main d'Œuvre



Source: (C) Enquete\_Opinion\_Industrie\_3T\_2017.pdf



Les Prévisions des prix

Source: (C) Enquete\_Opinion\_Industrie\_3T\_2017.pdf

Klaus Röder 06/12/2018 Pour le prochain trimestre, a la fin du 3eme trimestre de 2017, les chefs d'entreprises du secteur public prévoient une hausse de la production, de la demande, des prix de ventes et une stabilité des effectifs. Ceux du privé prévoient une baisse de l'activité, une stabilité de la demande. La plupart des enquêtés du secteur public envisage de bonnes perspectives de leurs trésorerie pour les trois prochains mois.

### 3.4. Comment arriver à des prévisions ?

- 1. Les enquêtes d'opinion
- 2. Les prévisions à la base de données du passée

Les prévisions économétriques à la base de données:

- Par Projection simple
- Par Lissage
- Par Régression simple et Multiple
- Par l'Analyse factorielle (AF) et/ou l'Analyse des Composantes Principales (ACP)
- Par les Techniques spéciales de combinaison de la Régression et l'ACP (a suivre)

Industrie : l'activité industrielle secteur privé trimestriel :



Exercices: Prévisions de la production de l'Industrie par:

- Par Projection simple
- Par Lissage
- Par Régression simple et Multiple
- Par l'Analyse factorielle (AF) et/ou l'Analyse des Composantes Principales (ACP)

Klaus Röder 06/12/2018 • Par les Techniques spéciales de combinaison de de la Régression et l'ACP



Source : Prévisions de la TC de la production de l'Industrie par: Lissage (1a. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/TC\_Ind\_Tri +ProjLiss(4Tr))



Prévisions du PIB de la production de l'Industrie par Régression Simple : Industrie expliqué par Hydrocarbures (1a. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/PIBInd\_Tri +ProjRegrSim)



Prévisions du PIB de la production de l'Industrie par Régression Multiple : Industrie expliqué par Hydrocarbures et Agriculture (1a. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBInd\_Tri +ProjRegrMulAgri)

Klaus Röder 06/12/2018



Prévisions du PIB de la production de l'Industrie par Régression Multiple : Industrie expliqué par Hydrocarbures et BTPH (1a. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/PIBInd\_Tri +ProjRegrMulBTPH)



Prévisions du PIB la production de l'Industrie par: l'Analyse factorielle (AF) et/ou l'Analyse des Composantes Principales (ACP) (1.J\_1a. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBInd\_Tri +ProjACP2Comp)



Evidemment les prévisions et résultats changent fortement et selon des objectives de la prévision peuvent être interprétés des manières différents et controverse.

### 4. Les Travaux en Groupe

Choisissez un group parmi les secteurs suivant le document (B) Le rapport de la Conjoncture du CNES actuel (Conjoncture-rapport-global CNES S12015.pdf) Le choix sera naturel selon votre spécialisation professionelle

Agriculture

Importation (Consommation du lait) Exportation Industrie Le secteur de l'énergie Les hydrocarbures La production industrielle manufacturière ISMMEE MAT. CONST. Céramique et verre Chimie, caoutchouc, plastiques Agro-alimentaires, tabacs, allumettes Textile, bonneterie, confection Bois, liège, papier et imprimé Industries diverses Les mines et carrières

Klaus Röder 06/12/2018

### Energie

Lecture de l'Evolution mensuelle et semestrielle des prix du pétrole Evolution des exportations de pétrole en valeurs Production et exportation de pétrole en volume Infrastructures Travaux publics: Habitat et urbanisme : Ressources en eau (Hydraulique) Services Commerce interne Transport Climat social/ Volet en vironnemental

Pour faciliter les travaux en groupe en propose un Schéma de Modélisation des Secteurs : Boites à outils conjoncturels

Ce schéma doit faciliter les travaux préparatoires pour utiliser les connaissances acquises

# 4.1. Les Schéma de Modélisation des Secteurs : Boites à outils conjoncturels

Etapes	Différentes dimensions	
A. Etapes pré	liminaire	
	Définition, mesures statistiques pertinentes des variables /Données disponibles/	A.1
	Connaissances théoriques/ Spécificités en Algérie : Eléments structurels/ conjoncturels	A.2
	Contenus autres rapports / Contenu rapport précédent du CNES aussi internationales	A.3
B. Analyse de	la conjoncture	
	Dernières mesures, Taux s et évolution descriptive récente depuis le dernier taux/graphiques	B.1
	Impact, multiplicateur / Manifestations d'excès, d'insuffisance, d'instabilité	B.2
	Les goulots d'étranglement de l'activité//Mécanismes déséquilibrants	B.3
C. Cadrage ma	acroéconomique	
	Variables en amont ; les raisons	C.1
	Investissement et programmes public, Dépenses publiques ; Législation, Instruments de politique économique	C.2
	Croissance /Décroissance/Variables en aval / Retombées sociales	C.3
D. Prévision à	court terme	
	Indices précurseurs	D.1
	Technique de prévision : Modélisation XLSTAT et autres	D.2
	Les Prévision et Modélisation pour les secteurs	D.3

### 4.2. Constitution des groupes

Trois ou Quatre groupes de travail devront être constitués; Ils sont relatifs aux thèmes des exercices de conjoncture. Chacun des thèmes sera traité par un groupe de trois à cinq apprenants. Il convient de préciser que le thème de la croissance globale de l'Algérie concerne tous les apprenants. De ce fait, nous avons un thème central et 3/4 groupes spécifiques.

Contenu des documents « Boites à outils conjoncturels »

Chaque groupe est invité à remettre un document regroupant toutes les dimensions à analyser. Il s'agit de disposer de la liste des instruments d'analyse conjoncturelle qu'il faut enrichir d'une manière progressive en vue d'obtenir d'une boite à outil la plus exhaustive possible dans un délai raisonnable.

Les contenus peuvent varier d'un document à un autre. Toutefois, les parties suivantes sont nécessaires

- Rédiger une note introductive expliquant l'intérêt du thème traité d'un point de vue théorique économique et d'un point de vue pratique en l'illustrant comme une dimension importante de la réalité économique et sociale de l'Algérie
- Mener une analyse conjoncturelle préliminaire aussi bien qualitative que quantitative essentiellement descriptive concernant l'évolution récente des grandeurs. Cette phase doit être complétée et argumentée par les principaux indicateurs (indicateurs avancés, informations du tableau de bord, etc.)
- Proposer une modélisation uni ou multi variée, avec une présentation synoptique de la méthodologie adoptée, en insistant sur l'interprétation des résultats empiriques obtenus. Cette modélisation doit être suivie de quelques prévisions sur le court terme des grandeurs traitées.

Utilisez l'approche du groupe pour poursuivre

### 4.3. Les travaux et Exercices du cours en groupe

Proposition des exercices par groupe pour arriver à des prévisions et l'analyse des secteurs :

- 1) Prévisions de la TC de la production du secteur choisi par: Projection simple linéaires
- 2) Prévisions de la TC de la production du secteur choisi par: Lissage (Moyennes mobiles)
- 3) Prévisions du PIB de la production du secteur choisi par Régression Simple :
- 4) Prévisions du PIB de la production du secteur choisi par Régression Multiple :

5) Prévisions du PIB de la production du secteur choisi par: l'Analyse factorielle (AF) et/ou l'Analyse des Composantes Principales (ACP)

Plus tard en proposera und technique de combinaisons des techniques 3-5 pour enrichir les connaissances de l'analyse et faciliter encore raisonnement et prévisions de court terme basé sur les analyses quantitatives des données

### 4.4. Préparations et Présentations des travaux en groupes :

Tandis que les étapes A et B sont plutôt verbal / écrite, les étapes C. et D. sont basés sur les exercices mentionnées. Utilisez les exercices pour les indicateurs du secteur du choix de votre groupe. .

**Pour A :**.Commencez avec la définition des statistiques pertinentes et une collection variables /Données disponibles

Pour les secteurs Agriculture//Industrie/Energie/Infrastructures/Services/ (ou les sous-secteurs) utilisez principalement les données du Fichier :

#### 1.a et 1b.Fichier-..J\_ONS\_PIB ALGERIE\_KR...xlsx

.. comme ils-contiennent les données nationales, annuelles et trimestrielles.

Si vous ne trouvez les données nécessaires cherchez dans les fichiers des donnes internationales (ex. Banque Mondial) :

#### 2.Algérie\_BM\_World\_Development\_Indicators180212.xlsx

Ce fichier ne contient que les données annuelles. Utilisez pour faciliter la compréhension les noms des séries traduits (Plage: Data\_fr)

Pour les autres secteurs comme le PIB et les Hydrocarbures choisissez les fichiers :

3.Algérie\_et\_Int\_GDP\_Data\_Extract\_From\_World\_Development\_Indicators180212.xlsx 4.AlgerieEtMonde\_UtilisationPétrole\_Data\_Extract\_From\_World\_Development\_Indicator s180212.xlsx

**Suite A et B:** Remplissez:Connaissances théoriques/Spécificités en Algérie ; Eléments structurels/ conjoncturels ; Contenus autres rapports / Contenu rapport précédent du CNES aussi internationales

Regardez aussi les rapports d'ONS :

- C Enquete\_Opinion\_Industrie\_3T\_17.pdf
- D Enquete\_Opinion\_Commerce\_3t\_16.pdf

Même si vous ne rappelez de tous les détails de la modélisation conjoncturelle, des méthodes statistiques et l'utilisation de XLSTAT consultez les documents :

F. et G. Manuel Statistique partiellement traduit:

(HandbookStats\_partiel\_fr\_KR171029.docx et Formation SPRING\_A7\_Manuel Statistique171020.pptx)

H. Introduction au logiciel XLSTAT (Formation SPRING\_A7\_XLSTAT\_171109.pptx

**Pour C et D :** Commencez avec une présentations des variables sélectionnés pour votre secteur. Si possible reprenez la référence au rapport du CNES, mais aussi aux autres publications comme lesquels du KOF.

Expliquez la signification des variables dans le contexte politique, économique et social. Expliquez les techniques de prévisions sans entrer dans les détails techniques, les lecteurs veulent savoir principalement du développement des secteurs et ne sont pas intéressés d'une introduction statistique ou économétrique.

Expliquez les résultats des prévisions sans cacher les incertitudes des toutes prévisions, mais préciser que vous utilisez des informations factuelles et expliquez ces méthodes dans la mesure où elles peuvent intéresser ou être utile pour les lecteurs des rapports du CNES.

# 5. La Régression PCR et PLS avec XLSTAT

Utilisez ce module pour modéliser et prédire les valeurs d'une ou plusieurs variables quantitatives en fonction d'une combinaison linéaire d'une ou plusieurs variables explicatives quantitatives et/ou qualitatives en vous affranchissant des contraintes de la régression linéaire pour ce qui concerne la distribution des variables et le nombre de variables que l'on peut inclure.

### 5.1. Description

Les trois méthodes de régression auxquelles ce module donne accès ont pour propriétés communes de générer des modèles mettant en jeu des combinaisons linéaires de variables explicatives. La différence entre les trois méthodes vient essentiellement de la façon dont sont traitées les structures de corrélations entre les variables.

### **Régression PCR**

La régression PCR (*Principal Components Regression*) ou régression sur composantes principales comprend trois étapes : on réalise d'abord une ACP (Analyse en Composantes Principales) sur le tableau des variables explicatives, puis on effectue une régression OLS (*Optimization Least Squares*) sur les composantes retenues, puis on calcule les paramètres du modèle sur les variables d'origine.

L'ACP permet de passer d'un tableau X comprenant *n* observations décrites par *p* variables à un tableau S de *n* observations décrites par *q* composantes, où *q* est inférieur ou égal à *p* et tel que (S'S) est inversible. Une sélection supplémentaire peut être effectuée de telle sorte que seuls les *r* composantes les plus corrélées avec la variable Y soient gardées pour la régression OLS. On obtient alors le tableau R.

Le calcul de la régression OLS s'effectue sur le tableau R. On obtient alors les paramètres correspondant à chacun des *r* facteurs. Afin de palier le problème d'interprétation des paramètres ainsi obtenus, XLSTAT effectue automatiquement les calculs nécessaires pour obtenir les paramètres et les intervalles de confiance pour les variables de départ.

### **Régression PLS**

Cette méthode est rapide, efficace et optimale pour un critère de minimisation des covariances bien maîtrisé. Son utilisation est recommandée dans le cas où un grand nombre de variables explicatives est utilisé, ou lorsqu'il y a de fortes colinéarités entre les variables.

L'idée de la régression PLS (*Partial Least Squares*) est de créer à partir d'un tableau de n observations décrites par p variables, un ensemble de *h* composantes avec h < p. La méthode de construction des composantes diffère de celle de l'ACP, et présente l'avantage de bien s'accommoder de la présence de données manquantes. La détermination du nombre de composantes à retenir est en général fondée sur un critère mettant en jeu une validation croisée. L'utilisateur peut aussi fixer lui-même le nombre de composantes à retenir. On distingue souvent la PLS1 de la méthode PLS2. La PLS1 concerne le cas où il y a une seule variable dépendante, la PLS2 celui où il y a plusieurs variables dépendantes. Les algorithmes utilisés dans XLSTAT sont tels que la PLS1 est un cas particulier de la PLS2. La distinction ne sera donc pas faite ici.

Dans le cas des méthodes OLS et PCR, si l'on doit calculer les modèles pour plusieurs variables dépendantes, le calcul des modèles consiste en une simple boucle sur les colonnes du tableau des variables dépendantes. Dans le cas de la régression PLS, la structure de covariance du tableau des variables dépendantes influe aussi sur les calculs.

L'équation du modèle de la régression PLS est donnée par

$$Y = T_k C'_k + E_k$$
  
=  $X W_k^* C'_k + E_k$   
=  $X W_k \left( P'_k W_k \right)^{-1} C'_k + E_k$ 

où Y est la matrice des variables dépendantes, X celle des variables explicatives, et où  $T_h$ ,  $C_h$ ,  $W^*_h$ ,  $W_h$  et  $P_h$ , sont des matrices générées par l'algorithme PLS, et où  $E_h$  est la matrice des résidus.

La matrice B des coefficients de régression de Y sur X en utilisant h composantes générées par l'algorithme de régression PLS est donc définie par :

$$B = W_k \left( P'_k W_k \right)^{-1} C'_k$$

Remarque : il s'agit donc comme en régression OLS ou PCR d'un modèle linéaire.

Remarques :

Les trois méthodes donnent le même résultat si le nombre de composantes issues de l'ACP (en régression PCR) ou de la PLS (régression PLS) est égal au nombre de variables explicatives sélectionnées.

En régression PLS, les composantes sont créées de fait de telle sorte qu'elles expliquent au mieux Y, alors qu'en PCR elles sont au départ créées uniquement en fonction de X. XLSTAT permet de corriger partiellement ce désavantage en proposant de sélectionner les composantes les plus corrélées avec Y.

Analyse discriminante PLS :

L'analyse discriminante PLS (PLS-DA) permet d'utiliser la méthode PLS pour expliquer et prédire l'appartenance d'individus à plusieurs classes, sur la base de variables explicatives quantitatives ou qualitatives. Pour cela, PLS-DA utilise l'algorithme PLS2 en prenant comme variable dépendante le tableau disjonctif complet obtenu à partir de la variable de classification qualitative.

L'analyse discriminante PLS permet d'effectuer une discrimination en utilisant les propriétés de l'algorithme PLS. Elle peut ainsi s'appliquer au cas de jeux de données avec peu d'observations et beaucoup de variables explicatives. Les composantes PLS sont utilisées et les mêmes options que dans le cas de la régression PLS sont disponibles. Elle permet aussi de n'utiliser que les données disponibles dans le cas de données manquantes et s'adapte très bien au cas de multi-colinéarité entre les variables explicatives.

On obtient donc autant de modèles que de modalités de la variable dépendante. Une observation est associée à la classe pour laquelle la valeur de l'équation du modèle est maximale.

Soit une variable dépendante à K modalités  $a_1, ..., a_K$ . Pour chaque modalité de la variable dépendante, on peut calculer pour chaque observation l'équation :

$$F(y_i, a_k) = b_0 + \sum_{j=1}^{p} b_j x_{ij}$$

Avec  $a_k$  une des modalités,  $b_0$  constante du modèle associé à la modalité  $a_k$  obtenu par régression PLS, p nombre de variable indépendantes et  $b_i$  coefficients de ce même modèle.

L'observation i est classée dans la classe k si :

$$k^* = \arg\max_k F(y_i, a_k)$$

L'analyse discriminante PLS offre une alternative intéressante à l'analyse discriminante linéaire classique.

Des sorties équivalentes à l'analyse discriminante de XLSTAT-Pro sont ainsi disponibles dans XLSTAT-PLS : la matrice de confusion, les valeurs prédites ainsi que la validation et la prédiction.

### 5.2. Boîte de dialogue

La boîte de dialogue est composée de plusieurs onglets correspondant aux différentes options disponibles tant pour la gestion des calculs que pour l'affichage des résultats. Vous trouverez cidessous le descriptif des différents éléments de la boîte de dialogue.

ok : cliquez sur ce bouton pour lancer les calculs.

- Annuler : cliquez sur ce bouton pour fermer la boîte de dialogue sans effectuer les calculs.
- Aide

: cliquez sur ce bouton pour afficher l'aide.



: cliquez sur ce bouton pour rétablir les options par défaut.



: cliquez sur ce bouton pour effacer les sélections de données.

: cliquez sur ce bouton pour changer la façon dont XLSTAT doit charger les données. Si la flèche est vers le bas, XLSTAT considère que les observations sont en lignes et les variables en colonnes. Si la flèche est vers la droite, XLSTAT considère que les variables sont en lignes et les observations en colonnes.

### Onglet **Général** :

### Y / Variables dépendantes :

Quantitatives/Qualitatives : sélectionnez la ou les variables dépendantes quantitatives (qualitatives dans le cas de la PLS-DA). Les données sélectionnées doivent être de type numérique (les données peuvent être nominales dans le cas de la PLS-DA). Si le libellé des variables a été sélectionné, veuillez vérifier que l'option « Libellés des variables » est activée.

### X / Variables explicatives :

**Quantitatives** : activez cette option si vous voulez inclure une ou plusieurs variables explicatives quantitatives. Sélectionnez alors la ou les variables correspondantes sur la feuille Excel. Les données sélectionnées doivent être de type numérique. Si le libellé des variables a été sélectionné, veuillez vérifier que l'option « Libellés des variables » est activée.

**Qualitatives** : activez cette option si vous voulez inclure une ou plusieurs variables explicatives qualitatives dans le modèle. Sélectionnez alors la ou les variables correspondantes sur la feuille Excel. Les données sélectionnées peuvent être de tout type, mais les données numériques sont

automatiquement considérées comme nominales. Si le libellé des variables a été sélectionné, veuillez vérifier que l'option « Libellés des variables » est activée.

Méthode : choisissez la méthode de régression à utiliser :

- **PLS-R** : activez cette option pour calculer une régression avec la méthode des moindres carrés partiels (Partial Least Squares).
- **PLS-DA** : activez cette option pour calculer une analyse discriminante avec la méthode des moindres carrés partiels (Partial Least Squares).
- PCR : activez cette option pour calculer une régression sur les composantes principales (Principal Components Regression).

**Plage** : si vous activez cette option, les résultats seront affichés à partir d'une cellule située dans une feuille existante. Vous devez alors sélectionner la cellule.

Feuille : activez cette option pour afficher les résultats dans une nouvelle feuille du classeur actif.

Classeur : activez cette option pour afficher les résultats dans un nouveau classeur.

Libellés des variables : activez cette option si la première ligne des données sélectionnées (variables dépendantes, explicatives, poids et libellés des observations) contient un libellé.

**Libellés des observations** : activez cette option si vous voulez utiliser des libellés d'observations disponibles sur une feuille Excel pour l'affichage des résultats. Si l'option « Libellés des variables » est activée, la première cellule de la sélection doit comprendre un en-tête. Si vous n'activez pas cette option, des libellés seront automatiquement créés (Obs1, Obs2, ...).

**Poids des observations** : activez cette option si vous voulez pondérer les observations. Si vous n'activez pas cette option, les poids seront tous considérés comme valant 1. Les poids doivent être impérativement supérieurs ou égaux à 0. Un poids de 2 est équivalent à répéter deux fois la même observation. Si un en-tête de colonne a été sélectionné, veuillez vérifier que l'option « Libellés des variables » est activée.

**Poids dans la régression** : cette option n'est active que pour les régressions PCR et OLS. Activez cette option si vous voulez effectuer une régression par les moindres carrés pondérés. Si vous n'activez pas cette option, les poids seront tous considérés comme valant 1. Les poids doivent être impérativement supérieurs ou égaux à 0. Si un en-tête de colonne a été sélectionné, veuillez vérifier que l'option « Libellés des variables » est activée.

Onglet **Options** :

Klaus Röder 06/12/2018 Options communes :

**Intervalle de confiance (%)** : entrez l'étendue en pourcentage de l'intervalle de confiance à utiliser pour les différents tests, et pour le calcul des intervalles de confiance autour des paramètres et des prédictions. Valeur par défaut : 95.

**Interactions / Niveau** : activez cette option pour inclure des interactions dans le modèle puis entrez le niveau maximum d'interaction (valeur comprise entre 1 et 4).

Options pour la régression PLS :

Conditions d'arrêt :

- Automatique : activez cette option que XLSTAT détermine automatiquement le nombre de composantes à retenir.
- Nombre fixé : activez cette option pour fixer le nombre de composantes à prendre en compte dans le modèle. La valeur par défaut est 2.
- Seuil Qi<sup>2</sup> : activez cette option pour fixer la valeur seuil du critère Qi<sup>2</sup> utilisée pour déterminer si l'apport d'une composante est significatif ou non pour l'une des variables dépendantes. La valeur par défaut est 0.0975 et correspond à 1-0.95<sup>2</sup>.
- Seuil Qi<sup>2</sup> (global) : activez cette option pour fixer la valeur seuil du critère Qi<sup>2</sup> utilisée pour déterminer si l'apport d'une composante est significatif ou non pour l'ensemble des variables dépendantes. La valeur par défaut est 0.0975 et correspond à 1-0.95<sup>2</sup>.
- Amélioration du Qi<sup>2</sup> : activez cette option pour fixer la valeur seuil du critère d'amélioration du Qi<sup>2</sup> utilisée pour déterminer si l'apport d'une composante est significatif ou non. La valeur par défaut est 0.05 et correspond à 5% d'amélioration. La valeur de ce critère est donnée par :

Amél.
$$Q^{2}(h) = \frac{Q^{2}(h) - Q^{2}(h-1)}{Q^{2}(h-1)}$$

• **Press minimum** : activez cette option pour que le nombre de composantes retenues corresponde au modèle donnant le coefficient de Press minimal.

#### X / Variables explicatives :

**Centrer** : activez cette option si vous voulez centrer les variables explicatives avant de commencer les calculs.

**Réduire** : activez cette option si vous voulez réduire les variables explicatives avant de commencer les calculs.

#### Validation croisée :

- Aucune : activez cette option pour ne pas effectuer de validation croisée. Dans ce cas les calculs seront plus rapides, mais les Q<sup>2</sup> et les intervalles de confiance ne pourront pas être calculés.
- Jackknife (LOO) : activez cette option pour effectuer une validation croisée basée sur un jackknife de type « leave one out ». Cette méthode n'est pas utilisable au-delà de 100 observations en raison de son coût en mémoire.

• **Jackknife** : activez cette option pour effectuer une validation croisée basée sur un jackknife avec répartition des données en k **groupes**. Un maximum de 100 groupes est accepté.

Options pour la régression PCR :

**ACP normée** : activez cette option pour effectuer une ACP sur la matrice de corrélation. Désactivez cette option pour effectuer une ACP sur la matrice de covariance.

**Filtrer les composantes** : vous pouvez activer l'une ou les deux options suivantes afin de réduire le nombre de composantes utilisées dans le modèle :

- % minimum : activez cette option puis saisissez le pourcentage minimum de la variabilité totale que doivent représenter les composantes sélectionnées.
- Nombre maximum : activez cette option pour fixer le nombre maximum de composantes à prendre en compte.

**Trier les composantes** : choisissez l'une des options suivantes afin de déterminer quel critère est utilisé pour trier les composantes avant que soient pris en compte les critères « % minimum » ou « Nombre maximum »:

- **Corrélations avec les Y** : activez cette option pour que la sélection des composantes se fasse après un tri décroissant suivant le carré du coefficient de corrélation (R<sup>2</sup>) entre la variable Y et les composantes. Cette option est recommandée.
- Valeurs propres : activez cette option pour que la sélection des composantes se fasse après un tri décroissant suivant les valeurs propres associées aux composantes.

**Constante fixée** : activez cette option pour fixer la constante du modèle de régression à une valeur que vous devez ensuite saisir (0 par défaut).

**Tolérance** : activez cette option pour permettre à l'algorithme de calcul de la régression OLS ne pas prendre en compte les variables qui seraient soit constantes soit trop corrélées avec d'autres variables déjà utilisées dans le modèle (0.0001 par défaut).

#### Onglet Validation :

**Validation** : activez cette option si vous souhaitez utiliser une partie des données sélectionnées pour valider le modèle.

Jeu de validation : choisissez l'une des options pour définir le mode de sélection des observations utilisées pour la validation :

- Aléatoire : les observations sont sélectionnées de manière aléatoire. Le « Nombre d'observations » doit alors être saisi.
- N dernières lignes : les N dernières observations sont sélectionnées pour la validation. Le « Nombre d'observations » N doit alors être saisi.
- N premières lignes : les N premières observations sont sélectionnées pour la validation. Le « Nombre d'observations » N doit alors être saisi.

• Variable de groupe : si vous choisissez cette option, vous devez ensuite sélection ner une variable indicatrice composée de 0 pour les observations à utiliser pour le calcul du modèle, et de 1 pour les observations à utiliser pour la validation du modèle.

#### Onglet **Prédiction** :

**Prédiction** : activez cette option si vous souhaitez sélectionner des données à utiliser en mode prédiction. Si vous activez cette option, vous devez veiller à ce que les données de prédiction soient organisées comme les données d'estimation : mêmes variables, même ordre dans les sélections. En revanche vous ne devez pas sélectionner de libellés de variables : la première ligne des sélections décrites ci-dessous doit être une ligne de données.

**Quantitatives** : activez cette option pour sélectionner la ou les variables quantitatives explicatives. La première ligne ne doit pas comprendre d'en-tête.

**Qualitatives** : activez cette option pour sélectionner la ou les variables qualitatives explicatives. La première ligne ne doit pas comprendre d'en-tête.

Libellés des observations : activez cette option si vous voulez utiliser des libellés d'observations disponibles sur une feuille Excel pour l'affichage des résultats. La première ligne ne doit pas comprendre d'en-tête. Si vous n'activez pas cette option, des libellés seront automatiquement créés (PredObs1, PredObs2, ...).

#### Onglet Données manquantes :

Ces options ne sont disponibles que dans le cas des régressions PCR et OLS. Pour la régression PLS, la gestion des données manquantes fait partie de la méthode.

Supprimer les observations : activez cette option pour supprimer les observations comportant des données manquantes.

Estimer les données manquantes : activez cette option pour estimer les données manquantes avant le début des calculs.

• Moyenne ou mode : activez cette option pour estimer les données manquantes en utilisant la moyenne (variables quantitatives) ou le mode (variables qualitatives) pour les variables correspondantes.

#### Onglet **Sorties** :

Options communes :

**Statistiques descriptives** : activez cette option pour afficher les statistiques descriptives pour l'ensemble des variables sélectionnées.

**Corrélations** : activez cette option pour afficher la matrice de corrélation pour les variables quantitatives (dépendantes et explicatives).

**Coefficients normalisés** : activez cette option pour afficher les paramètres normalisés du modèle (coefficients bêta).

**Equation** : activez cette option pour afficher explicitement l'équation du modèle.

**Prédictions et résidus** : activez cette option pour afficher les prédictions et les résidus pour l'ensemble des observations.

Options pour la régression PLS-R :

**Composantes t, u et u**<sup>~</sup> : activez cette option pour afficher les tableaux des composantes. Si cette option n'est pas activée, les graphiques correspondants ne sont pas affichés.

**Vecteurs c, w, w\* et p** : activez cette option pour afficher les tableaux des vecteurs générés par l'algorithme PLS. Si cette option n'est pas activée, les graphiques correspondants ne sont pas affichés.

**VIP** : activez cette option pour afficher le tableau et les graphiques correspondant aux *Variable Importance for the Projection* (VIP).

**Intervalles de confiance** : activez cette option pour calculer et afficher les intervalles de confiance autour des « coefficients standardisés ». Les calculs utilisent une méthode Jacknife.

Détection des valeurs extrêmes : activez cette option pour afficher le tableau et les graphiques des valeurs extrêmes.

Options pour la régression PLS-DA :

**Composantes t, u et u**<sup>~</sup> : activez cette option pour afficher les tableaux des composantes. Si cette option n'est pas activée, les graphiques correspondants ne sont pas affichés.

**Vecteurs c, w, w\* et p** : activez cette option pour afficher les tableaux des vecteurs générés par l'algorithme PLS. Si cette option n'est pas activée, les graphiques correspondants ne sont pas affichés.

**VIP** : activez cette option pour afficher le tableau et les graphiques correspondant aux *Variable Importance for the Projection* (VIP).

**Intervalles de confiance** : activez cette option pour calculer et afficher les intervalles de confiance autour des « coefficients standardisés ». Les calculs utilisent une méthode Jacknife.

Détection des valeurs extrêmes : activez cette option pour afficher le tableau et les graphiques des valeurs extrêmes.

**Matrice de confusion** : activez cette option pour afficher le tableau permettant de visualiser les nombres d'observations bien et mal classées pour chacune des classes.

Options pour la régression PCR-R : *Klaus Röder* 06/12/2018 **Coordonnées des variables** : activez cette option pour afficher les coordonnées des variables (*factor loadings* en anglais). Les coordonnées sont égales aux corrélations entre les composantes principales et les variables d'origine dans le cas d'une ACP normée.

**Corrélations Composantes/Variables** : activez cette option pour afficher les corrélations entre les composantes principales et les variables d'origine.

**Coordonnées des observations** : activez cette option pour afficher les coordonnées des observations (*factor scores* en anglais) dans le nouvel espace créé par l'ACP. Ces coordonnées sont ensuite utilisées dans l'étape OLS de la régression PCR.

Options pour les régressions PCR :

**Analyse de la variance** : activez cette option pour afficher le tableau d'analyse de la variance.**Type I SS** : activez cette option pour afficher le tableau de l'analyse de l'impact des variables de Type I (*Type I Sum of Squares*).**Type III SS** : activez cette option pour afficher le tableau de l'analyse de l'impact des variables de Type I (*Type III Sum of Squares*).**Analyse de la variance** : activez cette option pour afficher le tableau d'analyse de la variance.**Type I SS** : activez cette option pour afficher le tableau d'analyse de la variance.**Type I SS** : activez cette option pour afficher le tableau d'analyse de la variance.**Type I SS** : activez cette option pour afficher le tableau d'analyse de la variance de Type I (*Type II Sum of Squares*).**Type III SS** : activez cette option pour afficher le tableau de l'analyse de la variance de Type I (*Type II Sum of Squares*).**Type III SS** : activez cette option pour afficher le tableau de l'analyse de la variance de Type I (*Type III Sum of Squares*).**Prédictions ajustées** : activez cette option pour calculer et afficher les distances de Cook dans le tableau des prédictions et résidus.**D de Cook** : activez cette option pour calculer et afficher les distances de Cook dans le tableau des prédictions et résidus.**Press** : activez cette option pour calculer et afficher la statistique du Press (predicted residual error sum of squares).

Onglet Graphiques :

#### Options communes :

Graphiques de régression : activez cette option pour afficher les graphiques de régression :

- **Coefficients normalisés** : activez cette option pour afficher sur un graphique les paramètres normalisés du modèle avec leur intervalle de confiance.
- Prédictions et résidus : activez cette option pour afficher les graphiques suivants :

(1) Droite de régression : ce graphique n'est affiché que s'il n'y a qu'une seule variable explicative, et que cette variable est quantitative.

(2) Variable explicative versus résidus normalisés : ce graphique n'est affiché que s'il n'y a qu'une seule variable explicative, et que cette variable est quantitative.

(3) Variable dépendante versus résidus normalisés.

- (4) Prédictions pour la variable dépendante.
- (5) Graphique en bâtons des résidus normalisés.

• Intervalles de confiance : activez cette option pour afficher les intervalles de confiance sur les graphiques (1) et (4).

Options pour les régressions PLS et PCR :

**Graphiques de corrélations** : activez cette option pour afficher les graphiques mettant en jeu des corrélations entre des composantes et des variables initiales. Dans le cas de la PCR, activez cette option pour afficher le cercle des corrélations.

• Vecteurs : activez cette option pour afficher les variables d'origine sous forme de vecteurs.

**Graphiques des observations** : activez cette option pour afficher les graphiques de représentation des observations dans le nouvel espace.

• **Etiquettes** : activez cette option pour afficher les étiquettes des observations sur les graphiques. Le nombre d'étiquettes affichées peut être modulé à l'aide de l'option de filtrage.

**Biplots** : activez cette option pour afficher les graphiques de représentation simultanée des observations et des variables d'origine dans le nouvel espace.

- Vecteurs : activez cette option pour afficher les variables d'origine sous forme de vecteurs.
- **Etiquettes** : activez cette option pour afficher les étiquettes des observations sur les biplots. Le nombre d'étiquettes affichées peut être modulé à l'aide de l'option de filtrage.

**Etiquettes colorées** : activez cette option pour afficher les étiquettes de variables et d'observations de la même couleur que les points correspondants

Filtrer : activez cette option pour fixer le nombre de points affichés sur les graphiques :

• Aléatoire : les observations sont sélectionnées de manière aléatoire. Le « Nombre d'observations » doit alors être saisi. N premières lignes : les N premières observations sont affichées. Le « Nombre d'observations » N doit alors être saisi. N dernières lignes : les N dernières observations sont affichées. Le « Nombre d'observations » N doit alors être saisi. Variable de groupe : si vous choisissez cette option, vous devez ensuite sélectionner une variable indicatrice composée de 1 pour les observations à afficher, et de 1 pour les observations à ne pas afficher.

#### 5.3. Résultats

**Statistiques descriptives** : les tableaux de statistiques descriptives présentent pour toutes les variables sélectionnées des statistiques simples. Pour les variables dépendantes (en bleu) et les variables explicatives quantitatives, sont affichés le nombre d'observations, le nombre de données manquantes, le nombre de données non manquantes, la moyenne, et l'écart-type (non biaisé). Pour les variables explicatives qualitatives qualitatives sont affichés le nom des différentes modalités ainsi que leur fréquence respective.

**Matrice de corrélation** : ce tableau est affiché afin de vous permettre d'avoir un aperçu des corrélations entre les différentes variables sélectionnées.

Résultats de la régression PLS :

Klaus Röder 06/12/2018 Le premier tableau présente des indices de **qualité du modèle** sous forme de contribution cumulée des composantes aux indices :

L'indice  $Q^2$ cum est une mesure de l'apport global des *h* premières composantes à la qualité prédictive du modèle (et de ses sous-modèles s'il y a plusieurs variables dépendantes). L'indice  $Q^2$ cum(*h*) est défini par :

$$Q^{2}cum(h) = 1 - \prod_{j=1}^{k} \frac{\sum_{k=1}^{q} PRESS_{kj}}{\sum_{k=1}^{q} SCE_{k(j-1)}}$$

Cet indice fait intervenir un rapport des coefficients de PRESS (impliquant donc une validation croisée) et de la somme des carrés des erreurs (SCE) pour un modèle utilisant une composante de moins. La recherche du maximum de Q<sup>2</sup>cum revient donc à chercher le modèle le plus stable possible.

L'indice **R<sup>2</sup>Ycum** est la somme des coefficients de détermination entre les variables dépendantes et les h premières composantes. C'est donc une mesure du pouvoir explicatif des h premières composantes pour les variables dépendantes du modèle. L'indice **R<sup>2</sup>Xcum** est la somme des coefficients de détermination entre les variables explicatives et les h premières composantes. C'est donc une mesure du pouvoir explicatif des h premières composantes pour les variables explicatives et les h premières composantes. C'est donc une mesure du pouvoir explicatif des h premières composantes pour les variables explicatives du modèle. Un diagramme en bâtons est ensuite affiché afin de permettre une visualisation de l'évolution des trois indices en fonction du nombre de composantes. Si les R<sup>2</sup>Ycum et R<sup>2</sup>Xcum croissent nécessairement avec le nombre de composantes, ce n'est pas le cas pour Q<sup>2</sup>cum.

Le tableau suivant correspond à la **matrice de corrélation** des variables explicatives et dépendantes avec les composantes t et u. Un graphique permet ensuite de visualiser les corrélations avec les composantes t.

Le tableau des **vecteurs w** est ensuite affiché, suivi des tableaux des **vecteurs w\*** et des **vecteurs c** qui, comme il est montré dans la section « <u>Description</u> », interviennent directement dans le modèle. Si à h=2 correspond un modèle acceptable, il est démontré que la projection des vecteurs x sur les vecteurs y sur le **graphique des variables sur les w\*/c**, fournit une idée d'une part du signe dans le modèle des coefficients correspondant, d'autre part du poids relatif des variables de départ pour l'explication des variables dépendantes. Le tableau suivant correspond aux **coordonnées des observations** dans l'espace des composantes **t**. Le graphique est ensuite affiché. Si des observations de validation ont été sélectionnées, elles sont affichées sur ce graphique.

Le tableau des **coordonnées normalisées** est ensuite affiché. Ces coordonnées sont égales à la corrélation entre les variables indicatrices des observations et les composantes t. Ces coordonnées sont utilisées pour le graphique des **corrélations** qui suit, et qui permet la visualisation simultanée des observations, des variables dépendantes et des variables explicatives. Un exemple d'interprétation de ce graphique est disponible dans Tenenhaus (2003).

Le tableau suivant correspond aux **coordonnées des observations** dans l'espace des composantes **u** puis dans celui des composantes **u**<sup>~</sup>. Le graphique est ensuite affiché. Si des observations de validation ont été sélectionnées elles sont affichées sur ce graphique.

*Klaus Röder* 06/12/2018

Le tableau des **indices de qualité Q<sup>2</sup>** permet de voir comment les composantes contribuent à l'explication des variables dépendantes. Le tableau des **indices de qualité Q<sup>2</sup> cumulé** permet de mesurer la qualité associée à un espace de dimension croissante.

Les tableaux des **R<sup>2</sup> et des redondances** entre les variables de départ (dépendantes et explicatives) et les composantes t et u~ permettent de mesurer le pouvoir explicatif des composantes t et u~ tant au sens du R<sup>2</sup> qu'au sens de la redondance. La redondance entre un tableau X (n lignes et p variables) et une composante c est la part de la variance de X expliquée par c. On la définit comme la moyenne des carrés des coefficients de corrélation entre les

$$Rd(X,c) = \frac{1}{p} \sum_{j=1}^{p} R^{2}(x_{j},c)$$

variables et la composante :

Des redondances on peut alors déduire les VIP (Variable Importance for the Projection) qui mesurent l'importance d'une variable explicative pour la construction des composantes t. La VIP pour la variable explicative j et la composante h est définie par :

$$VIP_{kj} = \sqrt{\frac{p}{\sum_{i=1}^{k} Rd(Y, t_i)} \sum_{i=1}^{k} Rd(Y, t_i) w_{ij}^{2}}$$

Sur les graphiques des VIP (un diagramme en bâton par composante), une limite est tracée pour identifier les VIP supérieures à 0.8 ; il s'agit d'un seuil empirique proposé par Wold (1995) permettant d'identifier les variables fortement contributrices au modèle.

Le dernier suivant permet la **détection des valeurs extrêmes**. Les DModX (distances au modèle des observations dans l'espace des X) permettent d'identifier les valeurs anormales des variables explicatives, tandis que les DModY (distances au modèle des observations dans l'espace des Y) permettent d'identifier les valeurs anormales des variables dépendantes. Sur les graphiques correspondants sont affichés les seuils DCrit à partir desquels on peut considérer qu'une valeur de DMod est anormalement élevée. Les DCrit sont calculés en utilisant les valeurs seuil, classiquement calculées pour les box plots. La valeur de DModX pour la i<sup>ième</sup> observation est définie par :

$$DModX_i = \sqrt{\frac{n}{n-h-1} \frac{\sum_{j=1}^{p} e(X,t)_{ij}^2}{p-h}}$$

où les  $e(X,t)_{ij}$  (i=1 ...n) sont les résidus de la régression de X sur la j<sup>ième</sup> composante. La valeur de DModY pour la i<sup>ième</sup> observation est définie par :

*Klaus Röder* 06/12/2018

$$DModY_i = \sqrt{\frac{\sum_{j=1}^{q} e(Y, t)_{ij}^2}{q-h}}$$

où q est le nombre de variables dépendantes, et où les  $e(Y,t)_{ij}$  (i=1 ...n) sont les résidus de la régression de Y sur la j<sup>ième</sup> composante.

Le tableau qui suit présente les paramètres des modèles pour les différentes variables dépendantes, suivi des équations correspondantes si le nombre de variables explicatives est inférieur à 20.

Pour chacune des variables dépendantes est ensuite affichée une série de tableaux et graphiques.

**Statistiques d'ajustement** : dans ce tableau sont affichées les statistiques relatives à l'ajustement du modèle de régression, dont les définitions sont données dans la section consacréeLe tableau des **coefficients normalisés** (aussi appelés coefficients bêta) permet de comparer le poids relatif des variables dans le modèle. Pour le calcul des intervalles de confiance, dans le cadre de la PLS, les formules basées sur les hypothèses de normalité utilisées en régression OLS ne sont plus valables. Une méthode bootstrap proposée par Tenenhaus *et al* (2004) permet d'estimer les intervalles de confiance. Plus la valeur absolue d'un coefficient est élevée, plus le poids de la variable correspondante est important. Lorsque l'intervalle de confiance autour des coefficients normalisés comprend la valeur 0 (cela est facilement visible sur le graphique des coefficients normalisés), le poids d'une variable dans le modèle n'est pas significatif.

Dans le tableau des **prédictions et résidus** sont donnés pour chaque observation la valeur observée de la variable dépendante, la prédiction du modèle, les résidus et les intervalles de confiance. Deux types d'intervalles de confiance sont affichés : un intervalle de confiance autour de la moyenne (correspondant au cas où l'on ferait la prédiction pour un nombre infini d'observations avec un ensemble de valeurs données des variables explicatives) et un intervalle autour de la prédiction ponctuelle (correspondant au cas d'une prédiction isolée pour des valeurs données des variables explicatives). Le second intervalle est toujours plus grand que le premier, les aléas étant plus importants. Si des données de validation ont été sélectionnées, elles sont affichées en fin de tableau.

• Les **trois graphiques** affichés ensuite permettent de visualiser respectivement l'évolution des résidus en fonction de la variable dépendante, la distance entre les valeurs prédites et observées (pour un modèle idéal, les points seraient tous sur la bissectrice), le diagramme en bâtons des résidus.

Si vous avez sélectionné des données à utiliser pour calculer des prédictions, le tableau des prédictions est ensuite affiché.

Résultats spécifiques à l'analyse discriminante PLS :

Fonctions de classement : les fonctions de classement peuvent être utilisées pour déterminer à quelle classe doit être affectée une observation sur la base des valeurs prises pour les différentes variables explicatives. Une observation est affectée à la classe pour laquelle la fonction de classement est la plus élevée.

Classification a priori, a posteriori et scores de classement : dans ce tableau sont affichés pour chaque observation, sa classe d'appartenance définie par la variable dépendante, la classe d'appartenance telle que déduite des fonctions de classement et les valeurs des fonctions de classement associées à chacune des classes.

Matrice de confusion pour l'échantillon d'estimation : En utilisant les classifications a priori et a posteriori, on déduit la matrice de confusion, ainsi que le pourcentage global d'observations biens classées.

#### Résultats de la régression PCR :

La régression PCR requérant le calcul d'une Analyse en Composantes Principales, les résultats concernant cette dernière sont affichés.

**Valeurs propres** : les valeurs propres et le graphique (*scree plot*) correspondant sont affichés. Le nombre de valeurs propres est égal au nombre de valeurs propres non nulles. Si un filtrage a été demandé, il est appliqué au niveau de la régression elle-même.

Si les options de sorties correspondantes ont été activées, XLSTAT affiche ensuite les **coordonnées des variables** dans le nouvel espace, puis les corrélations entre les variables d'origine et les composantes dans le nouvel espace. Les **corrélations** sont égales aux coordonnées des variables dans le cas d'une ACP normée. Les **coordonnées des observations** dans le nouvel espace sont affichées dans un troisième tableau, et constituent les données utilisées ensuite pour la régression. Si des données de validation ont été sélectionnées, elles sont affichées en fin de tableau. Si l'option correspondante a été activée les biplots sont affichés. Si l'option de filtrage des composantes, s'appuyant sur les corrélations avec les variables dépendantes a été choisie, les composantes retenues pour la régression sont celles présentant les plus forts coefficients de détermination (R<sup>2</sup>) avec les variables dépendantes est alors affichée. Le nombre de composantes retenues dépend du nombre de valeurs propres et des options choisies (« % minimum » ou « Max composantes »).

Si l'option de filtrage des composantes s'appuyant sur les valeurs propres a été choisie, les composantes retenues pour la régression sont celles présentant les plus fortes valeurs propres. Le nombre de composantes retenues dépend du nombre de valeurs propres et des options choisies (« % minimum » ou « Max composantes »).

**Coefficients d'ajustement** : dans ce tableau sont affichées les statistiques relatives à l'ajustement du modèle de régression :

• **Observations** : le nombre d'observations prises en compte dans les calculs. Dans les formules présentées cidessous *n* désigne le nombre d'observations.

Klaus Röder 06/12/2018

- **Somme des poids** : la somme des poids des observations prises en compte dans les calculs. Dans les formules présentées ci-dessous *W* désigne la somme des poids.
- DDL : le nombre de degrés de liberté pour le modèle retenu (correspondant à la partie erreurs).
- **R<sup>2</sup>** : le coefficient de détermination du modèle. Ce coefficient, dont la valeur est comprise entre 0 et 1, n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par

$$\mathbb{R}^{2} = 1 - \frac{\sum_{i=1}^{n} w_{i} \left(y_{i} - \hat{y}_{i}\right)^{2}}{\sum_{i=1}^{n} w_{i} \left(y_{i} - \overline{y}\right)^{2}} \text{ avec } \overline{y} = \frac{1}{n} \sum_{i=1}^{n} w_{i} y_{i}$$

•

Le R<sup>2</sup> s'interprète comme la proportion

de la variabilité de la variable dépendante expliquée par le modèle. Plus le R<sup>2</sup> est proche de 1, meilleur est le modèle. L'inconvénient du R<sup>2</sup> est qu'il ne prend pas en compte le nombre de variables utilisées pour ajuster le modèle.

• **R<sup>2</sup> ajusté** : le coefficient de détermination ajusté du modèle. Le R<sup>2</sup> ajusté peut être négatif si le R<sup>2</sup> est voisin de zéro. Ce coefficient n'est affiché que si la constante du modèle n'est pas fixée par l'utilisateur. Sa valeur est définie par

$$\hat{\mathbb{R}}^2 = 1 - (1 - R^2) \frac{W - 1}{W - p - 1}$$

Le R<sup>2</sup> ajusté est une correction du R<sup>2</sup> qui permet de prendre en compte le nombre de variables utilisées dans le modèle.

$$MCE = \frac{1}{W - p^*} \sum_{i=1}^{n} w_i \left( y_i - \hat{y}_i \right)^2$$

MCE : la moyenne des carrés des erreurs (MCE) est définie par :

- **RMCE** : la racine de la moyenne des carrés des erreurs (RMCE) est la racine carrée de la MCE.
- MAPE : la Mean Absolute Percentage Error est calculée comme suit :

$$MAPE = \frac{100}{W} \sum_{i=1}^{N} w_i \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$DW = \frac{\sum_{i=2}^{n} \left[ (y_i - \hat{y}_i) - (y_{i-1} - \hat{y}_{i-1}) \right]^2}{\sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2}$$

**DW** : le coefficient de Durbin-Watson est défini par

Ce coefficient correspond au coefficient d'autocorrélation d'ordre 1 et permet de vérifier que les résidus du modèle ne sont pas autocorrélés, sachant que l'indépendance des résidus est l'une des hypothèses de base de la régression linéaire. L'utilisateur pourra se référer à une table des coefficients de Durbin-Watson pour vérifier si l'hypothèse d'indépendance des résidus est acceptable.

• **Cp** : le coefficient Cp de Mallows est défini par

$$Cp = \frac{SCE}{\hat{\sigma}} + 2p * -W$$

 où SCE est la somme du carré des erreurs pour le modèle avec p variables explicatives, et où correspond à l'estimateur de la variance des résidus pour le modèle comprenant toutes les variables explicatives. Plus le coefficient Cp est proche de p\* moins le modèle est biaisé.AIC : le critère d'information d'Akaike (Akaike's Information Criterion) est défini par

$$AIC = W \ln\left(\frac{SCE}{W}\right) + 2p *$$

• Ce critère proposé par Akaike (1973) dérive de la théorie de l'information, et s'appuie sur la mesure de Kullback et Leibler (1951). C'est un critère de sélection de modèles qui pénalise les modèles pour lesquels l'ajout de nouvelles variables explicatives n'apporte pas suffisamment d'information au modèle, l'information étant mesurée au travers de la SCE. On cherche à minimiser le critère AIC.**SBC** : le critère bayésien de Schwarz (Schwarz's Bayesian Criterion) est défini par

$$SBC = W \ln \left(\frac{SCE}{W}\right) + \ln (W) p^*$$

Ce critère proposé par Schwarz (1978) est proche du critère AIC, et comme ce dernier on cherche à le minimiser.

• PC : le critère de prédiction d'Amemiya (Amemiya's Prediction Criterion) est défini par laus Röder

$$PC = \frac{(1-R^2)(W+p^*)}{W-p^*}$$

Ce critère proposé par Amemiya (1980) permet comme le R<sup>2</sup> ajusté de tenir compte de la parcimonie du modèle.**Press** : la statistique du Press (predicted residual error sum of squares) n'est affichée que si l'option correspondante a été activée dans la boîte de dialogue. Elle est définie par

$$Press = \sum_{i=1}^{n} w_i \left( y_i - \hat{y}_{i(-i)} \right)^2$$

où  $\hat{\mathcal{P}}_{(-i)}$  est la prédiction pour l'observation i lorsque cette dernière n'est pas utilisée pour l'estimation des paramètres. On obtient alors

Press RMCE = 
$$\sqrt{\frac{\text{Press}}{W - p^*}}$$

Le Press RMCE peut alors être comparé au RMCE. Une différence importante entre les deux indique que le modèle est sensible à la présence ou absence de certaines observations dans le modèle.

Le **tableau d'analyse de la variance** permet d'évaluer le pouvoir explicatif des variables explicatives. Dans le cas où la constante du modèle n'est pas fixée à une valeur donnée, le pouvoir explicatif est évalué en comparant l'ajustement (au sens des moindres carrés) du modèle final avec l'ajustement du modèle rudimentaire composé d'une constante égale à la moyenne de la variable dépendante. Dans le cas où la constante du modèle est fixée, la comparaison est faite par rapport au modèle pour lequel la variable dépendante serait égale à la constante fixée.

Dans le cas d'une régression PCR, le premier tableau des **paramètres du modèle** correspond aux paramètres du modèle s'appuyant sur les composantes principales sélectionnées. Ce tableau est difficilement interprétable. Pour cette raison, une transformation est opérée afin d'obtenir les **paramètres du modèle** correspondant aux variables d'origine. Ce dernier tableau est obtenu directement dans le cas d'une régression OLS. Dans ce tableau, pour la constante du modèle et pour variable sont affichés l'estimation du paramètre, l'écart-type correspondant, le t de Student et la probabilité associée, ainsi que l'intervalle de confiance.

L'équation du modèle est ensuite affichée pour faciliter la lecture ou la réutilisation du modèle.

Le tableau des **coefficients normalisés** (aussi appelés coefficients bêta) permet de comparer le poids relatif des variables. Plus la valeur absolue d'un coefficient est élevée, plus le poids de la variable correspondante est important. Lorsque l'intervalle de confiance autour des coefficients normalisés comprend la valeur 0 (cela est facilement visible sur le graphique des coefficients normalisés), le poids d'une variable dans le modèle n'est pas significatif.

Dans le tableau des **prédictions et résidus** sont donnés pour chaque observation, son poids, la valeur de la variable explicative qualitative s'il n'y en a qu'une, la valeur observée de la variable dépendante, la prédiction du modèle, les résidus, les intervalles de confiance, ainsi que la prédiction ajustée et le D de Cook si les options correspondantes ont été activées dans la boîte de dialogue. Deux types d'intervalles de confiance sont affichés : un intervalle de confiance autour de la moyenne (correspondant au cas où l'on ferait la prédiction pour un nombre infini d'observations avec un ensemble de valeurs données des variables explicatives) et un intervalle autour de la prédiction ponctuelle (correspondant au cas d'une prédiction isolée pour des valeurs données des variables explicatives). Le second intervalle est toujours plus grand que le premier, les aléas étant plus importants. Si des données de validation ont été sélectionnées, elles sont affichées en fin de tableau.

Les **graphiques** qui suivent permettent de visualiser les résultats mentionnés ci-dessus. S'il n'y a qu'une seule variable explicative dans le modèle, le premier graphique affiché permet de visualiser les valeurs observées, la droite de régression et les deux types d'intervalles de confiance autour des prévisions. Le second graphique permet quant à lui de visualiser les résidus normalisés en fonction de la variable explicative. En principe, les résidus doivent être distribués de manière aléatoire autour de l'axe des abscisses. L'observation d'une tendance ou d'une forme révèlerait un problème au niveau du modèle.

Les **trois graphiques** affichés ensuite permettent de visualiser respectivement l'évolution des résidus normalisés en fonction de la variable dépendante, la distance entre les prédictions et les observations (pour un modèle idéal, les points seraient tous sur la bissectrice), et les résidus normalisés sous la forme d'un diagramme en bâtons. Ce dernier graphique permet de rapidement voir si un nombre anormal de données sort de l'intervalle ]-2, 2[ sachant que ce dernier, sous hypothèse de normalité, doit contenir environ 95% des données.

Si vous avez sélectionné des données à utiliser pour calculer des **prédictions sur de nouvelles observations**, le tableau correspondant est ensuite affiché.

# 5.4. Exemple d'un tutoriel XLS: Régression PLS (moindres carrées partiels)

### Jeu de données pour réaliser une régression par les moindres carrées partiels - PLS

Ce tutoriel utilise des données présentées dans l'article [Tenenhaus, M., Pagès, J., Ambroisine L. and & Guinot, C. (2005). PLS methodology for studying relationships between hedonic judgements and product characteristics. Food Quality an Preference. 16, 4, pp 315-325]. Par ailleurs il est recommandé aux utilisateurs francophones de XLSTAT d'acquérir le livre très complet de Michel Tenenhaus sur la régression PLS :

Les données utilisées dans l'article correspondent à 6 jus d'orange évalués suivant 16 descripteurs physico-chimiques, et notés par 96 juges.

L'utilisation de la régression PLS va permettre d'obtenir une cartographie simultanée des juges, des descripteurs et des produits, puis d'analyser juge par juge quelles caractéristiques soustendent les préférences exprimées. Un classeur Excel comprenant à la fois les données utilisées dans cet exemple et les résultats obtenus peut être téléchargé en cliquant sur le bouton ci-dessous : <u>Télécharger les données</u> Paramétrer une régression par les moindres carrées partiels - PLS

Pour activer la boîte de dialogue de la régression PLS, lancez XLSTAT, puis sélectionnez la commande **Régression PLS** de la barre d'outils **Modélisation des données**.



Une fois que vous avez cliqué sur le bouton, la boîte de dialogue apparaît. Sélectionnez au niveau des **Variables dépendantes** (les "Y" du modèle), les notes données par les juges.

Ce sont en effet les données que l'on veut expliquer au travers des Variables explicatives quantitatives (les "X" du modèle) que sont les descripteurs des jus d'orange.

Puis cliquez sur l'onglet **Options** et fixer le nombre de composantes à 4 dans les conditions d'arrêt.

Enfin, dans l'onglet **Graphiques**, l'option **Etiquettes colorées** a été activée afin de rendre les graphiques plus lisibles, et l'option **vecteurs** a été désactivée afin de ne pas alourdir inutilement les graphiques.

Régression PLS	X
Général Options Validation Prédiction Y / Variables dépendantes : Quantitatives : Data!\$R\$1:\$DI\$7	Données manq.   Sorties   Graphiques   C Plage :
X / Variables explicatives : Quantitatives : Data!\$B\$1:\$Q\$7 Qualitatives : 	<ul> <li>Libellés des variables</li> <li>Libellés des observations :</li> <li>Data!\$A\$1:\$A\$7</li> <li>Poids des observations :</li> </ul>
" ∕ ◆ ⊘ ∎	OK Annuler Aide

Les calculs, extrêmement rapides, commencent lorsque l'on a appuyé sur OK.

L'affichage des résultats est interrompu pour vous permettre de choisir les axes de représentation des résultats.



Il suffit de cliquer sur **Terminer** pour que les graphiques concernent uniquement les deux premiers axes.

L'affichage des résultats prend ensuite quelques secondes du fait de la présence de 96 variables dépendantes et des très nombreux tableaux et graphiques affichés.

#### Interpréter les résultats d'une régression par les moindres carrées partiels - PLS

Après quelques statistiques de base sur les différentes variables sélectionnées (les variables explicatives sont en noir, et les variables dépendantes en bleu) et la matrice de corrélations correspondante, les résultats propres à la régression PLS sont affichés.

Le premier tableau et le graphique correspondant permettent de visualiser la qualité de la régression PLS en fonction du nombre de composantes retenues.

*Klaus Röder* 06/12/2018



L'indice Q<sup>2</sup> cumulé est une mesure globale de la qualité de l'ajustement et de la qualité prédictive des 96 modèles. XLSTAT a automatiquement retenu 4 composantes. On voit que l'indice Q<sup>2</sup> reste faible. Cela suggère que la qualité de l'ajustement peut être très variable en fonction des juges.

Les R<sup>2</sup>Y cum et R<sup>2</sup>X cum qui correspondent aux corrélations entre les composantes et les variables de départ sont proches de un dès la quatrième composante, ce qui indique que les composantes sont à la fois bien représentatives des X et des Y.

Le premier graphique des corrélations permet de visualiser sur les deux premières composantes générées par la régression PLS les corrélations entre X et les Y du modèle.



On note que pour quelques juges au centre du graphique les corrélations sont très faibles. En se reportant au tableau, on voit par exemple que le juge 54 n'est lié qu'à la composante t4, globalement peu corrélée avec les variables explicatives.

En ce qui concerne les variables explicatives, on note que seule la vitamine C est mal représentée sur le graphique. Cette dernière est donc globalement faiblement explicative des préférence des juges, ce qui se comprend bien puisque elle n'a pas d'implication directe sur les critères gustatifs qui influencent les juges.

On notera les fortes corrélations positives entre le fructose et le glucose, entre les pH, et la corrélation bien entendu négative entre les pH d'une part, et l'acide citrique et le titre d'autre part. On voit apparaître sur ce graphique une grande différence de préférences entre les juges.

La carte superposant les variables dépendantes sur les vecteurs c et les variables explicatives sur les vecteurs w\* permet de visualiser la relation globale entre les variables, sachant que les w\* sont représentatifs du poids des variables dans les modèles.



Si l'on projète une variable explicative sur le vecteur d'une variable dépendante (les vecteurs sont affichés s'il y a moins de 50 variables dépendantes et si l'option vecteurs est activée), on a une idée de son poids dans le modèle concernant cette même variable dépendante.

Un tableau donne les coordonnées des produits sur les composantes t. Le graphique correspondant sur les axes t1 et t2 est ensuite affiché. On notera que les produits sont bien distingués.



Un dernier graphique de corrélations permet de superposer les produits sur le graphique des corrélations précédemment affiché. Dans la légende, "Obs" a été remplacé manuellement par "Jus", en modifiant l'intitulé de la série dans la barre de formule Excel, après avoir cliqué sur la série pour la sélectionner. Comme presque toujours avec XLSTAT, les graphiques sont des graphiques Excel totalement modifiables.



Klaus Röder 06/12/2018 Dans leur article précédemment cité, Tenenhaus et al. interprètent ce graphique en détail. Ils en déduisent notamment l'existence de 4 groupes bien identifiés de juges, sur lesquels il conseillent de réaliser des analyses séparées. Ils obtiennent alors des Q<sup>2</sup> et R<sup>2</sup> cumulés plus élevés. Pour le premier groupe identifié ils obtiennent un R<sup>2</sup>Y de 0.63 au lieu du 0.53 observé ici.

Deux tableaux fournissant des résultats pour les composantes u et u~ sont ensuite affichés. Un graphique permet de visualiser les observations (ici les jus d'orange) dans l'espace des u~.

Les tableaux qui suivent permettent de voir l'évolution des indices Q<sup>2</sup> et Q<sup>2</sup> cumulé en fonction du nombre de composantes. pour l'ensemble des variables dépendantes. On remarque que pour plusieurs variables que le maximum du Q<sup>2</sup> est atteint avec une ou deux composantes (voir par exemple J5, J6, J7).

Une série de tableaux présentant les R<sup>2</sup> pour chacune des variables d'entrée avec les composantes t est ensuite optionnellement affichée. L'option n'étant pas activée par défaut, les tableaux ne sont pas pris en compte dans ce tutoriel.

Le tableau suivant présente les VIP (Variable Importance for the Projection) pour chacune des variables explicatives, sur chacun des modèles avec un nombre croissant de composantes. Cela permet d'identifier rapidement quelles sont les variables explicatives les plus importantes sur l'ensemble des modèles. Sur la première composante, la Vitamine C, le Pouvoir Sucrant, l'Intensité Odeur et l'Intensité Goût apparaissent comme étant peu influentes.



Le tableau des paramètres des modèles correspondant à chacune des variables dépendantes est ensuite affiché. Les équations sont ensuite affichées afin de faciliter une éventuelle utilisateur ultérieure.

Pour chaque modèle sont ensuite affichés le tableau des coefficients d'ajustement, le tableau des coefficients normalisés (correspondant aux coefficients bêta de la régression linéaire classique) et enfin le tableau des prédictions et résidus. L'analyse du modèle correspondant au juge J1 nous permet de conclure à une bonne qualité du modèle, le R<sup>2</sup> vallant 0.88. Cependant, le nombre de degrés de liberté est faible (DDL =1), et on risque un sur-ajustement du modèle. Cela tend à être confirmé par le fait que pour l'ensemble des coefficients normalisés, les intervalles de confiance comprennent la valeur 0. Etant donné que le Q<sup>2</sup> cumulé diminue dès la troisième composante (cf tableau des Q<sup>2</sup>), il est fort probable qu'une qualité similaire aurait été atteinte avec seulement deux composantes.

Nous avons donc réalisé une nouvelle régression PLS, en ne sélectionnant que le juge 1, et en forçant XLSTAT à ne prendre en compte que deux composantes. Les résultats sont disponibles sur la feuille PLS2. On obtient alors des résultats plus riches. Le graphique ci-dessous correspond aux coefficients normalisés pour le modèle avec 2 composantes.



On voit ici que seuls les coefficients de l'intensité d'odeur et la typicité d'odeur sont significatifs. Le tableau des prédictions et résidus permet de vérifier que les notes du juge 1 sont très bien reproduites par le juge 1.

Enfin le tableau des DModX et DModY et les graphiques correspondant permettent d'identifier rapidement des observations qui constitueraient d'éventuels outliers (valeurs extrêmes). Ici, aucune valeur anormale n'a été détectée. Toutes les valeurs sont en effet inférieures aux DCritX et DCritY.

### 5.5. Exercices préparés

La régression PLS est une technique récente qui généralise et combine-les caractéristiques de l'analyse sur composantes principales et de la régression multiple.

Elle est particulièrement utile quand on a besoin de prédire un ensemble de variables dépendantes partir d'un ensemble très grand de variables explicatives (prédicteurs) qui peuvent être très fortement corrélées entre elles.

Le modèle de régression multiple a été étendu de plusieurs façons afin de s'adapter aux problèmes d'analyse de données plus sophistiqués. Il sert donc de base pour de nombreuses méthodes multivariées comme l'analyse discriminante, la régression sur composantes principales (PCR) et autres.

En statistique, la régression en composantes principales (PCR) est une technique d'analyse de régression basée sur l'analyse en composantes principales (ACP). En règle générale, il considère la régression du résultat (également connu comme la réponse à la variable dépendante) sur un ensemble de prédicteurs, ou variables explicatives, ou variables indépendantes basé sur un modèle de régression linéaire standard, **mais utilise ACP pour estimer les coefficients de régression inconnus dans le modèle.** Commençons avec la régression en composantes principales (PCR) :



Prévisions du PIB de la production de l'Industrie par: Projection PCR simple : PIB d'Industrie expliqué par PIB Agri (1b. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBInd\_Tri+ProjPCR1(2Ser)) :



Prévisions du PIB de la production de l'Industrie par: Projection PCR simple : PIB d'Industrie expliqué par PIB Agri, Industries diverses, Eau et Energie, Mines et Carrières. Comparaisons des Prévisions pour les trimestres futures (1b. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBInd\_Tri+ProjPCR2Comp(2+5Ser))

La régression PLS pour le modèle linéaire s'applique au champ d'application où l'on possède peu d'observations sur des variables très corrélées et en très grand nombre

La régression sur composantes principales (PCR) et la régression PLS diffèrent dans la méthode utilisée pour extraire les facteurs de scores. En résumé, la régression sur composantes principales produit une matrice de poids W reflétant la structure de covariances entre les variables prédictives alors que la régression PLS produit une matrice de poids W reflétant les structures de covariance entre les prédicteurs et les réponses.

L'intérêt de la méthode comparée à la régression sur composantes principales (PCR), réside dans le fait que les composantes PLS sur les X, notées t, sont calculées "dans le même temps" que des régressions partielles sont exécutées. Cette simultanéité leur confère un meilleur pouvoir prédictif que celles de la PCR.

Le nombre total d'axes ou de composantes (la dimension du modèle) est le super-paramètre de la méthode. Lors de l'exécution de l'algorithme, il est utile de voir évoluer, à chaque nouvel axe construit, la reconstruction de la variance des prédicteurs et des réponses. La stratégie consiste à s'arrêter lorsque le pourcentage de la variance de X est suffisamment grand, pour un gain faible dans le pourcentage de variance de Y. Ce critère est donc basé sur rajustement des données.



Prévisions du PIB de la production de l'Industrie par: Projection PLS avec 2 composantes et 11 Séries (3 Dépendantes et 8 explicatives) Les PIB d'Industrie (Industrie, .Ind.-Agri-Ali et Ind. divers) expliqués par les séries diverses, (1b. J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBInd\_Tri+ProjPLS(2Comp+11S)



Klaus Röder 06/12/2018 Page 57

Prévision du PIB du Commerce et des Services par: Projection PLS avec 4 composantes et 15 Séries (4 Dépendantes et 11 explicatives) expliqués par Séries diverses et comparé avec le même exercice sans une variable particulière (1.b\_J\_ONS\_PIB ALGERIE\_KR180219.xlsx/ PIBComSer\_Tri+ProjPLS2(4Comp)+PIBComSer\_ProjPLS2(4C-AjPISB))

En a changé les exercices pour calculer les Prévisions du Commerce et Services (Voyez aussi le dernier document trimestrielle du ONS: (D) Enquete\_Opinion\_Commerce\_3t\_16.pdf On a utilisé quatre composants mai une deuxième fois avec une amélioration du modèle par une modification des variable (en laissé tomber le variable «Ajustement PISB ». Voyez les changements légers des prévisions

Depuis la découverte de la méthode PLS, de nombreux chercheurs ont tenté de l'améliorer ou de l'étendre à d'autres champs d'applications proposent la PLS-DA (PLS Discriminant Analysis) une méthode généralisant PLS au contexte de la discrimination.

Dans l'analyse discriminante linéaire et l'analyse des variables canoniques étroitement apparentées, en a la matrice X divisé en un certain nombre de classes. L'objectif de l'analyse est de trouver des combinaisons linéaires des variables X (les fonctions discriminantes) entre classes, c'est-à-dire, ont des valeurs très différentes pour les classes. A condition que chaque classe soit serré et occupe un petit volume séparé dans l'espace X, on peut trouver un plan - un plan discriminant - dans lequel le les observations sont bien séparées selon classe. Avec de nombreuses variables X colinéaires, un PLS version PLS-DA est utile



Prévision du PIB du Commerce et plusieurs autres secteur par: Projection PLS-DA avec 4 composantes et 30 Séries (5 Dépendantes et 25 explicatives) expliqués par Séries diverses (1.b\_J\_ONS\_PIB ALGERIE\_KR180219.xlsx/PIBetComSecteurs\_ProjPLS-DA)

Klaus Röder 06/12/2018 Page 58

### 5.6. Les Exercices du cours en groupe

Utilisez l'approche au-dessus et les sélections des groupes pour poursuivre Proposition des exercices par groupe pour arriver à des prévisions et l'analyse des secteurs :

- 1) Prévisions du PIB de la production du secteur par Projection PCR simple ou Prévisions de s TC (choisissez un nombre adéquat des séries)
- Prévisions du PIB de la production du secteur par Projection PLS avec 2 composantes ou Prévisions de s TC, Variez le nombre des composantes (Choix fixe) avec le même choix des variables
- 3) Prévisions du PIB de la production du secteur par Projection PLS-DA ou Prévisions de s TC, Comparez les prévisions avec les prévisions des autres méthodes PCR/PLS

Pour l'utilisation des fichiers des données referez-vous au chapitre 4.4.

Expliquez la signification des variables dans le contexte politique, économique et social. Expliquez les techniques de prévisions sans entrer dans les détails techniques, les lecteurs veulent savoir principalement du développement des secteurs et ne sont pas intéressés d'une introduction statistique ou économétrique.

Expliquez les résultats des prévisions sans cacher les incertitudes des toutes prévisions, p.e. les marge de confiance, mais préciser que vous utilisez des informations factuelles et expliquez ces méthodes dans la mesure où elles peuvent intéresser ou être utile pour les lecteurs des rapports du CNES.

### Bibliographie :

- [1] Guide des indicateurs de marché Economiques et financiers Stefan Keller (Auteur), Julien Browaeys, 2008, Dunod
- [2] Déchiffrer les statistiques économiques et sociales Stéphanie Dupays, 2008, Dunod
- [3] Analyse des séries temporelles Application à l'économie et à la gestion Régis Bourbonnais , Michel Terraza, 2016, Dunod
- [4] KOF Centre de recherches conjoncturelles, Bulletin N° 107, mai 2017, https://www.kof.ethz.ch/fr/
- [5] ETH Zurich & KOF Index of Globalization, <u>http://globalization.kof.ethz.ch/</u>
- [6] Économétrie Cours et exercices corrigés, Régis Bourbonnais, 9e Edition, 2015, Dunod
- [7] Statistical Poverty-Analysis , Klaus Röder, 2009; <u>http://www.klaus-roeder.com/Ordner/PDFs/PovertyUserMan\_090507.pdf</u>
- [8] TURINFO (Influence of tourism on the economic development in 5 Brazil regions) , <u>http://www.klaus-</u>

```
roeder.com/4_Projekte/Trainings_Projects/1994_1997%20TURIN%20FO.html
```

 [9] Business Surveys and Forecasts (Industry, Trade, Commerce, Construction) for Paraguay, <u>http://www.klaus-</u>

```
roeder.com/Ordner/PDFs/1997InformeCeppro_GTZ_ES_1+2.pdf (español)
```

[10] Business Surveys and Forecasts (Commerce, Construction and Tourism) for Cabo Verde,- all Klaus Röder, 2000-2007, <u>http://www.klaus-</u>

# roeder.com/4\_Projekte/projekte.html

- [11] Results of IFTRAB 2004/2005:Comparing Labor Conditions in Formal and Informal Sector in Sofala, Mozambique, Klaus Röder, <u>http://www.klaus-</u>
- <u>roeder.com/Ordner/DocumentsWork/[19]Analysis%20Informal%20Sector061026.pdf</u> [12] Use of Water and Groups of Economical Strength in Maputo, Klaus Röder,
- 2007, <u>http://www.klaus-roeder.com/Ordner/DocumentsWork/[20]UseofWater-</u> <u>ResultadosAguaMaputoMatola070604.pdf</u>
- [13] Forecasting Economic Activity in Germany , Felix P. Hufner and Michael Schröder, « How Useful are Sentiment Indicators ? », 2£WDiscussion Paper No. 02-56 (Sep. 2002) <u>ftp://ftp.ZGW.de/pub/zew-docs/dp/dp0256.pdf</u>
- [14] Akaike H. (1973). Information Theory and the Extension of the Maximum Likelihood Principle. In: Second International Symposium on Information Theory. (Eds: V.N. Petrov and F. Csaki). Academiai Kiadó, Budapest. 267-281.
- [15] Amemiya T. (1980). Selection of regressors. International Economic Review, 21, 331-354.
- [16] Bastien P., Esposito Vinzi V. and Tenenhaus M. (2005). PLS generalised regression. Computational Statistics and Data Analysis, 48, 17-46.
- [17] Dempster A.P. (1969). Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading.
- [18] Kullback S. and Leibler R. A. (1951). On information and sufficiency. Annals of Mathematical Statistics, 22, 79-86.

- [19] Schwarz G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.
- [20] Tenenhaus M. (1998). La Régression PLS, Théorie et Pratique. Technip, Paris.
- [21] Tenenhaus M., Pagès J., Ambroisine L. and Guinot C. (2005). PLS methodology for studying relationships between hedonic judgements and product characteristics. Food Quality and Preference. 16, 4, 315-325.
- [22] Vancolen Séverine (2004) La régression PLS

# https://doc.rero.ch/record/7954/files/mem\_VancolenS.pdf

- [23] Wold, S., Martens H. and Wold H. (1983). The Multivariate Calibration Problem in Chemistry solved by the PLS Method. In: Ruhe A.and Kågström B. (eds.), Proceedings of the Conference on Matrix Pencils. Springer Verlag, Heidelberg. 286-293.
- [24] Wold S. (1995). PLS for Multivariate Linear Modelling. In: van de Waterbeemd H. (ed.), QSAR: Chemometric Methods in Molecular