

Désaisonnalisation par modélisation avec JDemetra+

Paris, le 2 avril 2015
Jean.palate@nbb.be



Plan

- ▶ JDemetra+
 - Objectifs, architectures, algorithmes
- ▶ Désaisonnalisation par modélisation
 - Modèles de référence (approche AMB, STS)
 - Description
 - Estimation
 - Comparaison
 - Modèles alternatifs
 - Intérêt de la modélisation
 - Analyse des révisions
 - Limites de la modélisation



1 Jdemetra+

un outil flexible pour
l'ajustement saisonnier

[et pour les algorithmes
connexes]

Projet Eurostat



JDemetra+ officially recommended as software for the seasonal adjustment of official statistics



1.1 JDemetra+. Objectifs

► Objectifs généraux

- Algorithmes utiles à la production/analyse des statistiques officielles
 - Séries chronologiques régulières (mensuelles → annuelles)
 - Algorithmes pour
 - ◆ Ajustement saisonnier, analyse des cycles
 - ◆ Benchmarking, temporal disaggregation
 - ◆ Modélisation (prévisions, détection d'anomalies...)
- Modules compatibles avec les infrastructures IT actuelles
 - Java, WEB services...
- Dessiné pour l'ensemble du processus statistique
 - De la recherche à la production de masse (→ flexible, performant)
- Maintenable à long terme
 - Solution "Open Source" (<https://github.com/jdemetra>)



▶ Objectifs liés à l'ajustement saisonnier

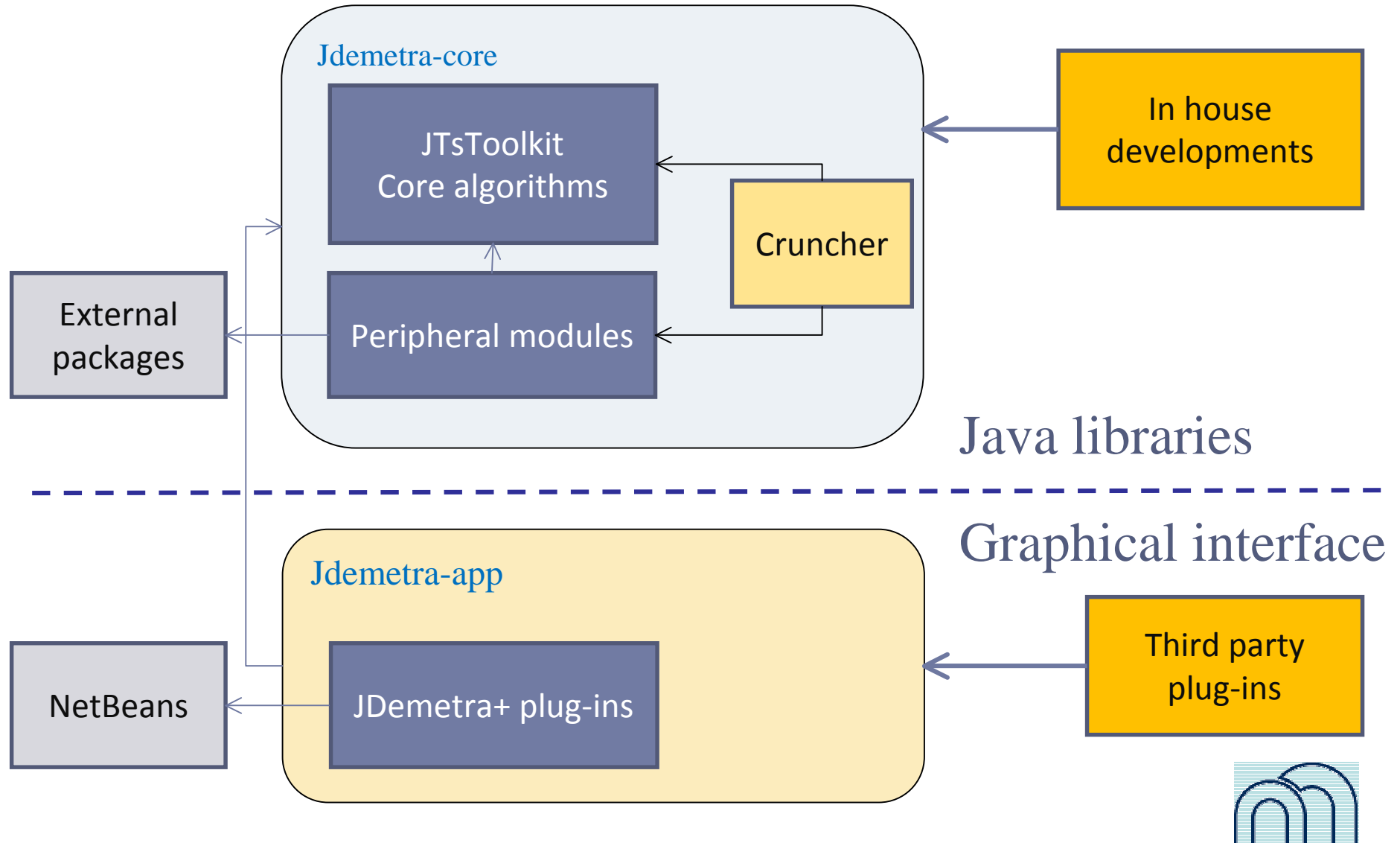
- Application des « guidelines » d'Eurostat
 - Tramo-Seats, X12-ARIMA [, modèles structurels...]
- Exploration de solutions alternatives et des problèmes connexes (benchmarking...)
- Implémentation en Java des algorithmes de référence
 - Modules OO flexibles et réutilisables

▶ Aspects organisationnels

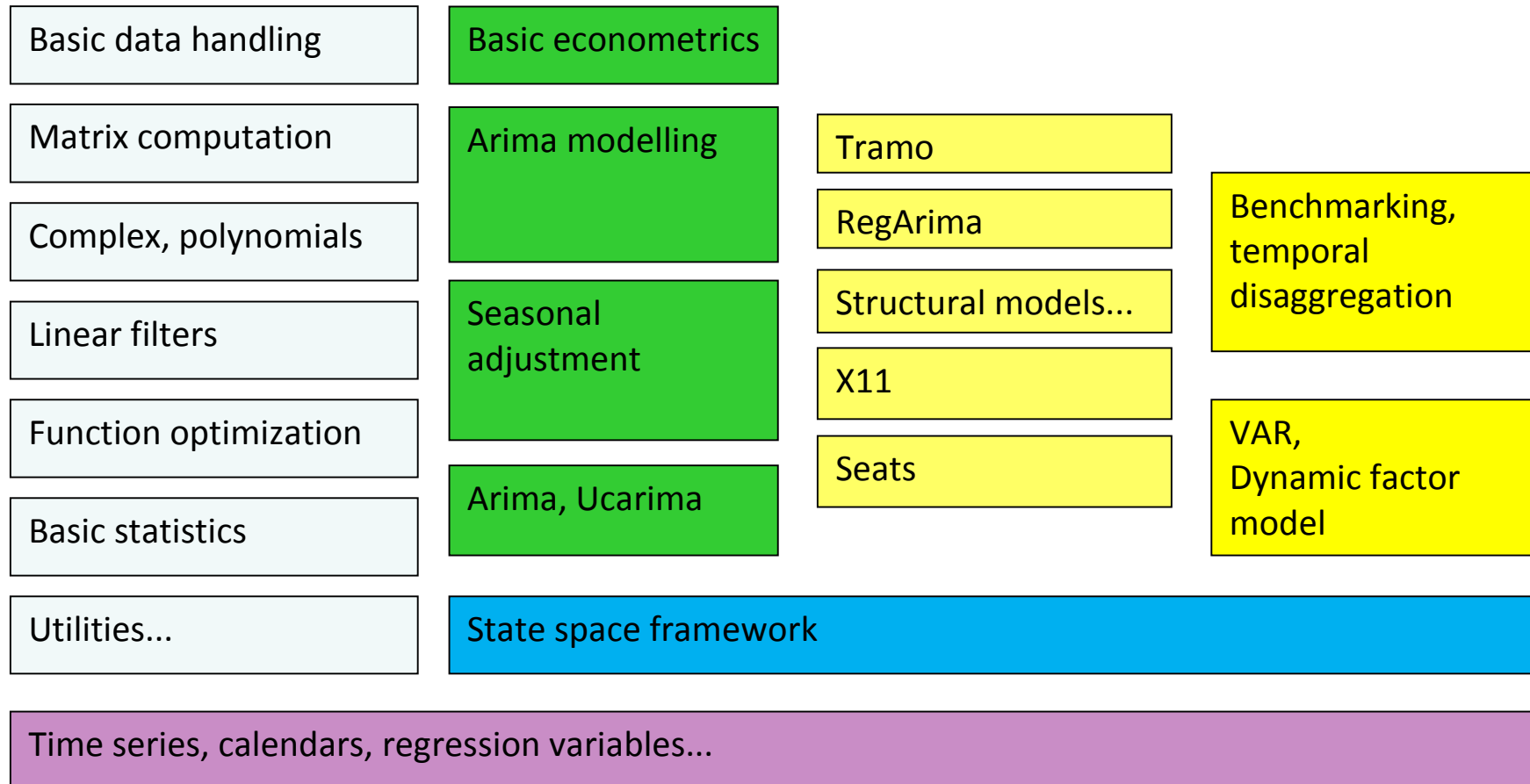
- Décisions: « Seasonal Adjustment Steering Group » (Eurostat/ECB)
- Développements: BNB et Bundesbank
- Support: « Seasonal adjustment Center of Excellence » (INSEE, ISTAT, ONS, STATEC)



1.2 Jdemetra+. Architecture



1.3 Jdemetra+. Bibliothèques algorithmiques

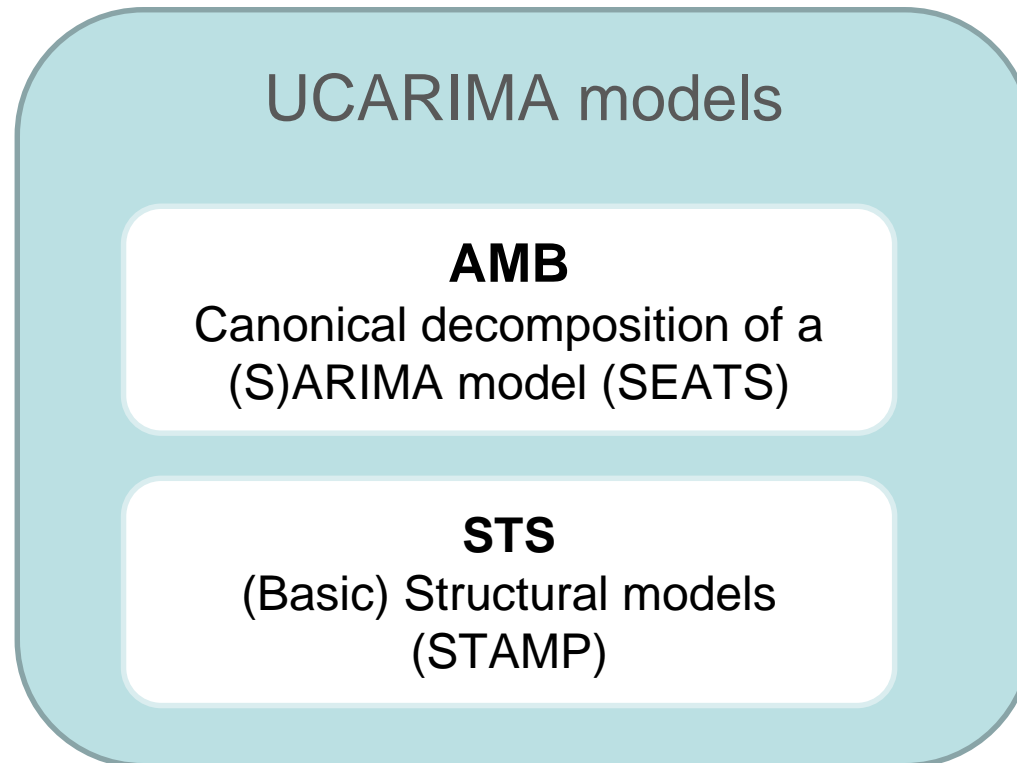


2 Désaisonnalisation par modélisation

- ▶ Principales motivations:
 - Comprendre les propriétés statistiques des méthodes utilisées.
 - Dériver les algorithmes qui correspondent à des hypothèses à priori.
- ▶ Avantage de JD+
 - Grande variété de modèles/algorithmes
- ▶ Plan
 - Modèles de référence (approche AMB/STS)
 - Méthodes d'estimation (WK/KF)
 - Comparaison AMB/STS
 - Extension des modèles standards



2.1 Modèles de référence. Description



2.11 Modèles UCARIMA

- ▶ « Somme » de modèles ARIMA indépendants
- ▶ Modèle simplifié « signal + bruit » : $y_t = s_t + n_t$

$$\begin{aligned}\Phi(B)y_t &= \Theta(B)\varepsilon_t, & \varepsilon_t &\sim N(0, v) \\ \Phi_s(B)s_t &= \Theta_s(B)\gamma_t, & \gamma_t &\sim N(0, v_s) \\ \Phi_n(B)n_t &= \Theta_n(B)\delta_t, & \delta_t &\sim N(0, v_n)\end{aligned}$$

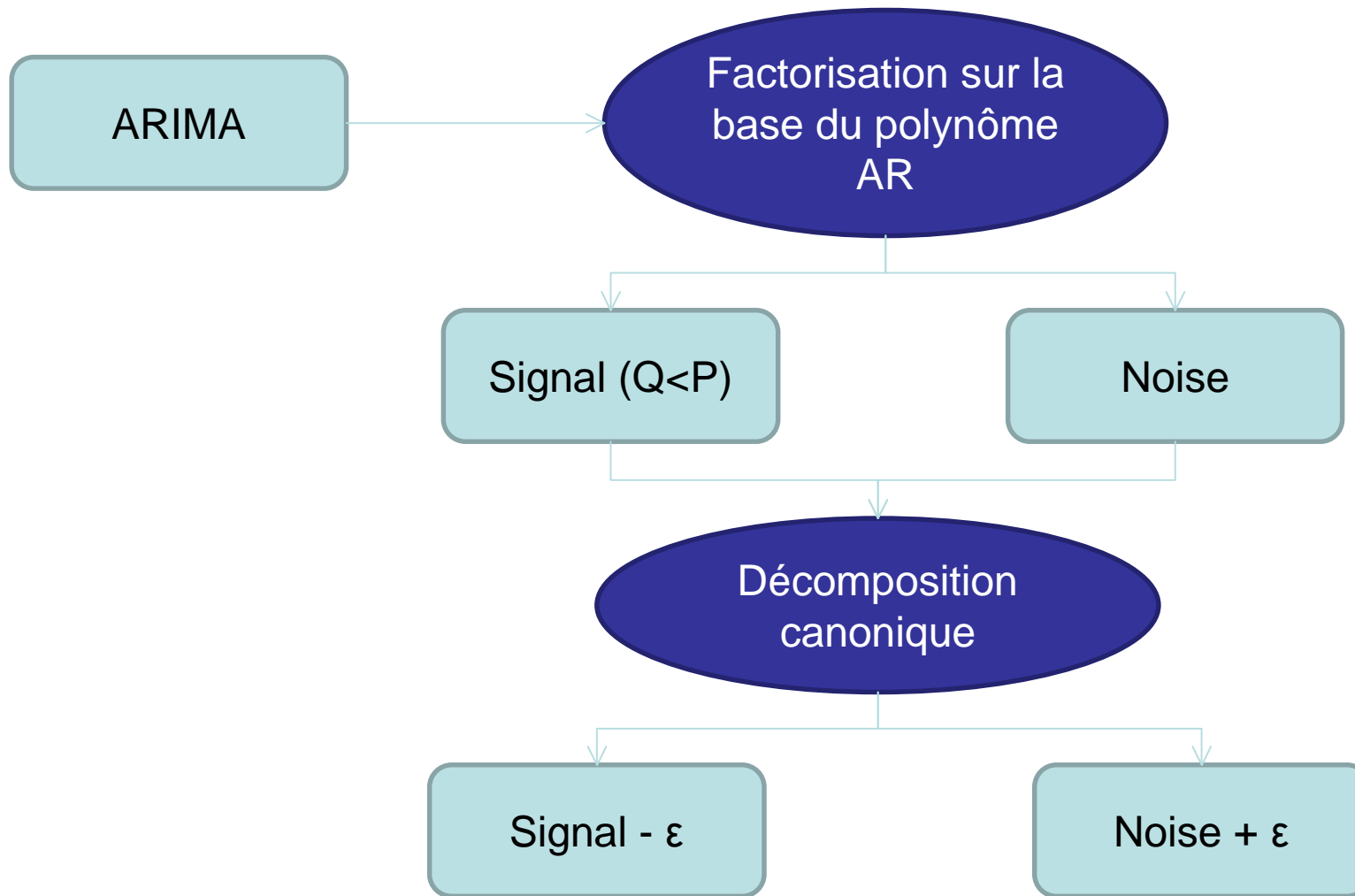
- ▶ Contraintes sur les polynômes:

$$\Phi_s(B)\Phi_n(B) = \Phi(B)$$

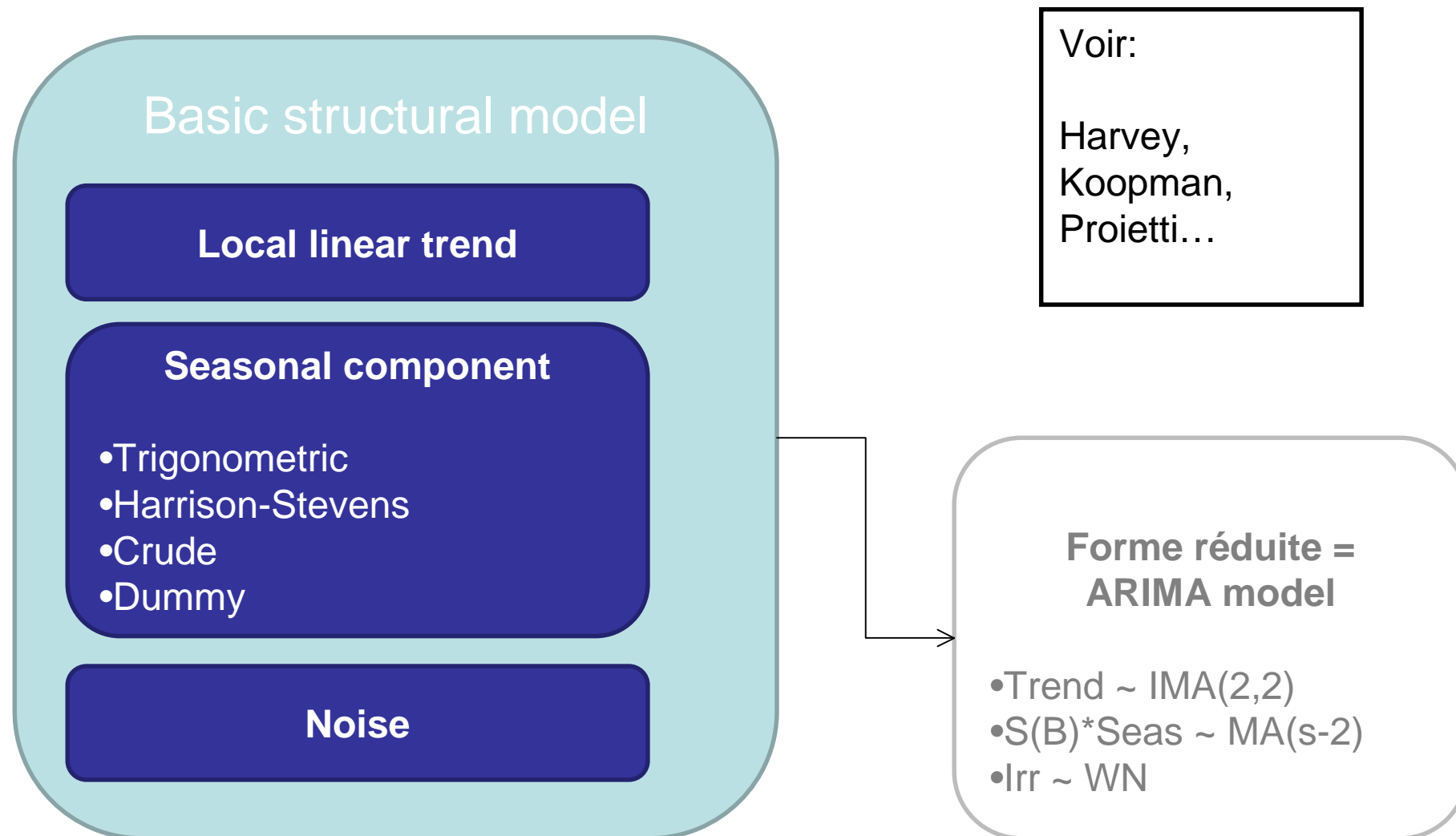
$$v_s\Phi_n(B)\Theta_s(B)\Phi_n(F)\Theta_s(F) + v_n\Phi_s(B)\Theta_n(B)\Phi_s(F)\Theta_n(F) = v\Theta(B)\Theta(F)$$



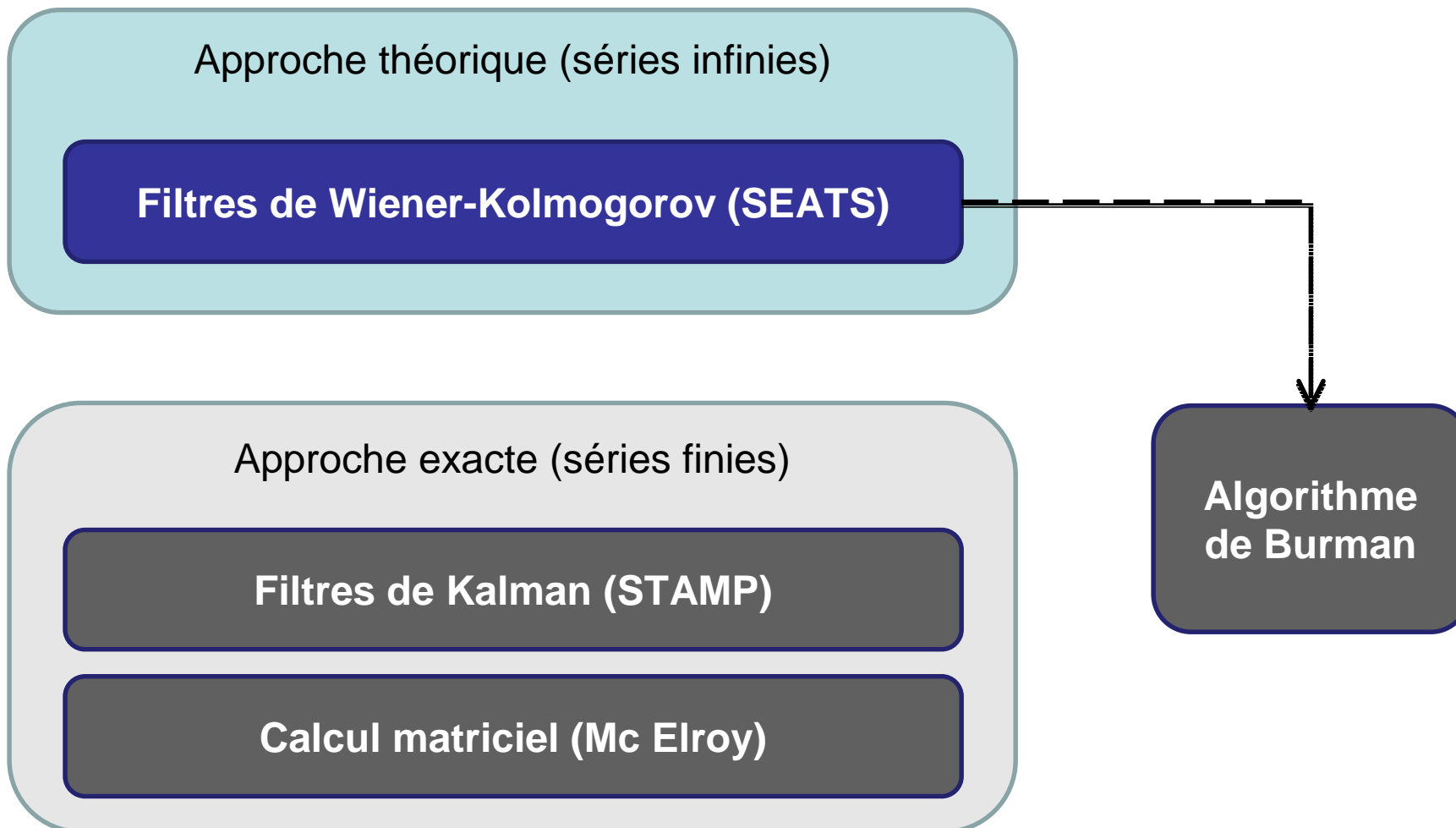
2.12 Arima model-based (AMB)



2.13 STS (dans JD+)



2.2 Méthodes d'estimation



2.21 Filtrés de Wiener-Kolmogorov (WK)

- ▶ Estimation des composantes:
 - Algorithme de Burman
 - Forecasts, backcasts sur quelques périodes + application des filtres
 - Solution simple, très performante, instable si modèle quasi-non inversible
- ▶ Propriétés du modèle:
 - Grille d'analyse développée dans SEATS (Maravall)
 - Approche théorique, basée sur une série infinie
 - Solution simple et performante
 - Approximation parfois médiocre si modèle quasi-non inversible
- ▶ Applicable uniquement à des modèles invariants dans le temps



► Estimateurs MMSE

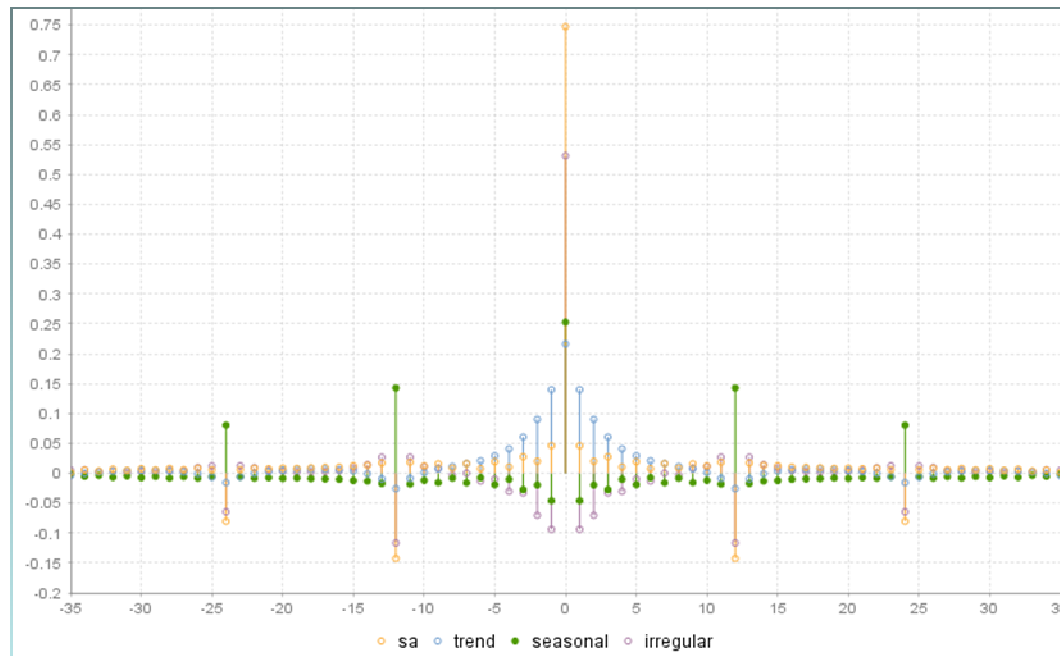
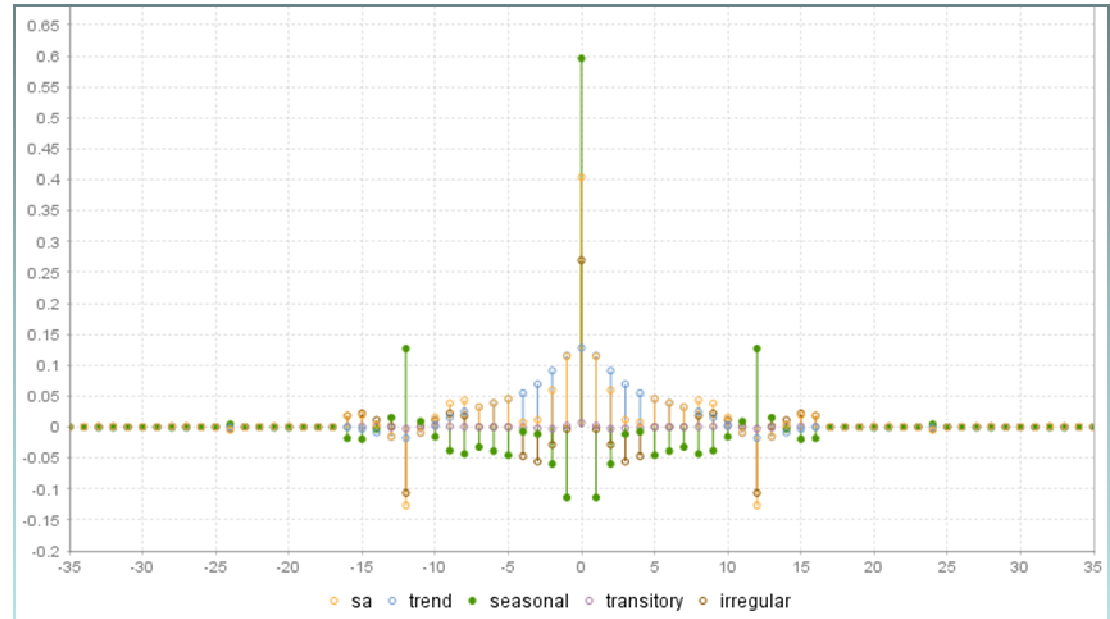
$$\hat{s}_t = \frac{v_s \frac{\Theta_s(B)\Theta_s(F)}{\Phi_s(B)\Phi_s(F)}}{v \frac{\Theta(B)\Theta(F)}{\Phi(B)\Phi(F)}} y_t$$

$$\hat{s}_t = \frac{v_s \Phi_n(B)\Theta_s(B)\Phi_n(F)\Theta_s(F)}{v\Theta(B)\Theta(F)} y_t$$

$$\hat{n}_t = \frac{v_n \Phi_s(B)\Theta_n(B)\Phi_s(F)\Theta_n(F)}{v\Theta(B)\Theta(F)} y_t$$



Décomposition canonique



Modèle structurel



- ▶ Expression en fonction des innovations du modèle agrégé:

$$\hat{S}_t = \frac{v_s \Theta_s(B) \Phi_n(F) \Theta_s(F)}{\Phi_s(B) \Theta(F)} \varepsilon_t$$

- ▶ PsiE-Weights (SEATS)

$$\hat{S}_t = \sum_{i=0}^{\infty} \xi_i^- \varepsilon_{t-i} + \sum_{i=1}^{\infty} \xi_i^+ \varepsilon_{t+i}$$



2.22 Filtrés de Kalman (KF)

- ▶ Initialisation des modèles non stationnaires via l'approche diffuse de Koopman
- ▶ Estimation des composantes
 - Kalman smoother
 - Disturbance smoother
- ▶ Propriétés du modèle
 - Estimations exactes (modifications/extension du modèle espace-état)
 - solution parfois complexe à mettre en œuvre
- ▶ Extension aux modèles variables dans le temps



► Modèles espace-état dans JD+

$$y_t = Z_t \alpha_t$$

$$\alpha_{t+1} = T_t \alpha_t + \mu_t, \mu_t \sim N(0, \sigma^2 V_t)$$

[+ conditions initiales habituelles]

► Modèle espace-état de la décomposition

$$y_t = [Z_s \quad Z_n] \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{st+1} \\ \alpha_{nt+1} \end{bmatrix} = \begin{bmatrix} T_s & 0 \\ 0 & T_n \end{bmatrix} \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Omega_s & 0 \\ 0 & \Omega_n \end{bmatrix}\right)$$



► Représentation de la composante saisonnière dans le modèle espace-état. Exemple: modèle Harrison-Stevens

- Représentation de type « random walk »

$$\gamma_t = X_t \delta_t, \quad X_t \text{ dummies saisonnières,}$$

$$\delta_{i,t+1} = \delta_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, v)$$

- Contrainte: $\sum_i \delta_{it} = 0 \rightarrow$ utilisation de “contrastes saisonniers”

$$\gamma_t = \tilde{X}_t \tilde{\delta}_t, \quad \tilde{X}_{it} = X_{it} - X_{st},$$

$$\tilde{\delta}_{t+1} = \tilde{\delta}_t + \mu_t, \quad \mu_t \sim N(0, \Omega), \quad \Omega = \frac{1}{s} \begin{bmatrix} s-1 & -1 & \dots & -1 \\ -1 & s-1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \dots & -1 & s-1 \end{bmatrix}$$

- Remarques:
 - Modèle identique pour des effets de jours ouvrables variant dans le temps
 - Généralisation possible à un modèle d'hétéroscédasticité saisonnière



► Représentation générale de la composante saisonnière dans le modèle espace-état STS de JD+ (West-Harrison)

$$\alpha_{\gamma,t+1} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \alpha_{\gamma,t} + \mu_{\gamma,t}, \quad \mu_{\gamma,t} \sim N(0, \Omega_t), \quad \Omega_t \text{ dépendant du modèle}$$

$$\gamma_t = [1 \quad 0 \quad \dots \quad 0] \alpha_{\gamma,t}$$

Remarque:

Avec cette représentation, on a toujours: $S(B)\gamma_t = \theta(q), \quad q \leq s - 2$

$$\theta_{hs} = 1 + .79B + .62B^2 + .47B^3 + .35B^4 + .25B^5 + .17B^6 + .10B^7 + .06B^8 + 0.03B^9 + 0.01B^{10}$$

$$\theta_{tri} = 1 + .74B + .63B^2 + .43B^3 + .36B^4 + .22B^5 + .18B^6 + .09B^7 + .07B^8 + 0.02B^9 + 0.02B^{10}$$

$$\theta_{crude} = 1 + .91B + .82B^2 + .73B^3 + .64B^4 + .55B^5 + .45B^6 + .36B^7 + .27B^8 + 0.18B^9 + 0.09B^{10}$$

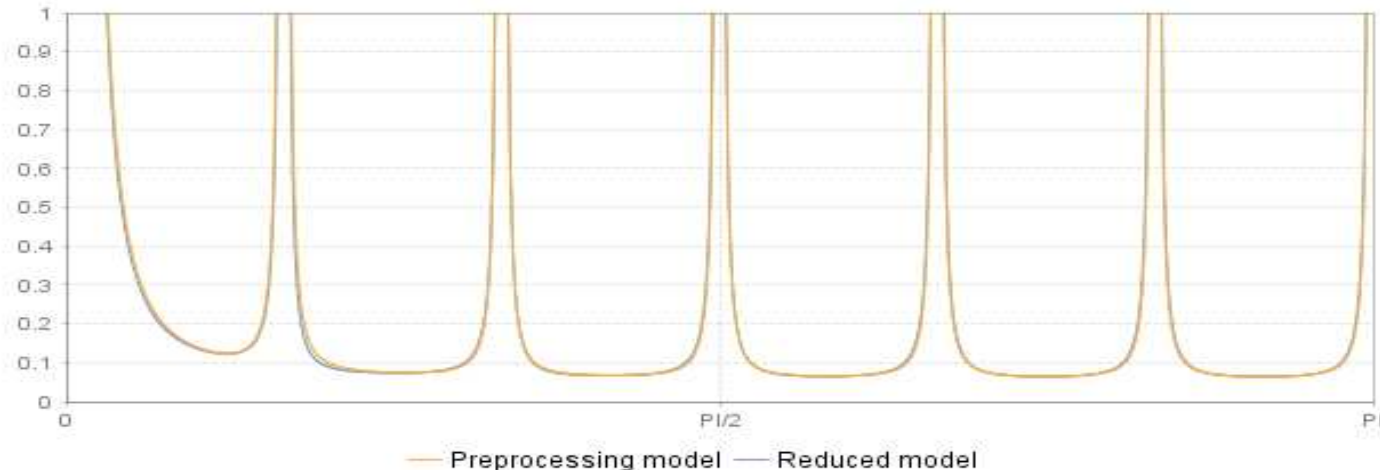
$$\theta_{dummy} = 1$$



2.3 Comparaison de AMB et STS

► Modélisation de la série

- Comparaison d'un Airline et d'un STS
(série INSEE 001654507)



Preprocessing model (0,1,1)(0,1,1)

Polynomials

regular MA: 1.00000 - 0.709555 B
seasonal MA: 1.00000 - 0.487508 S

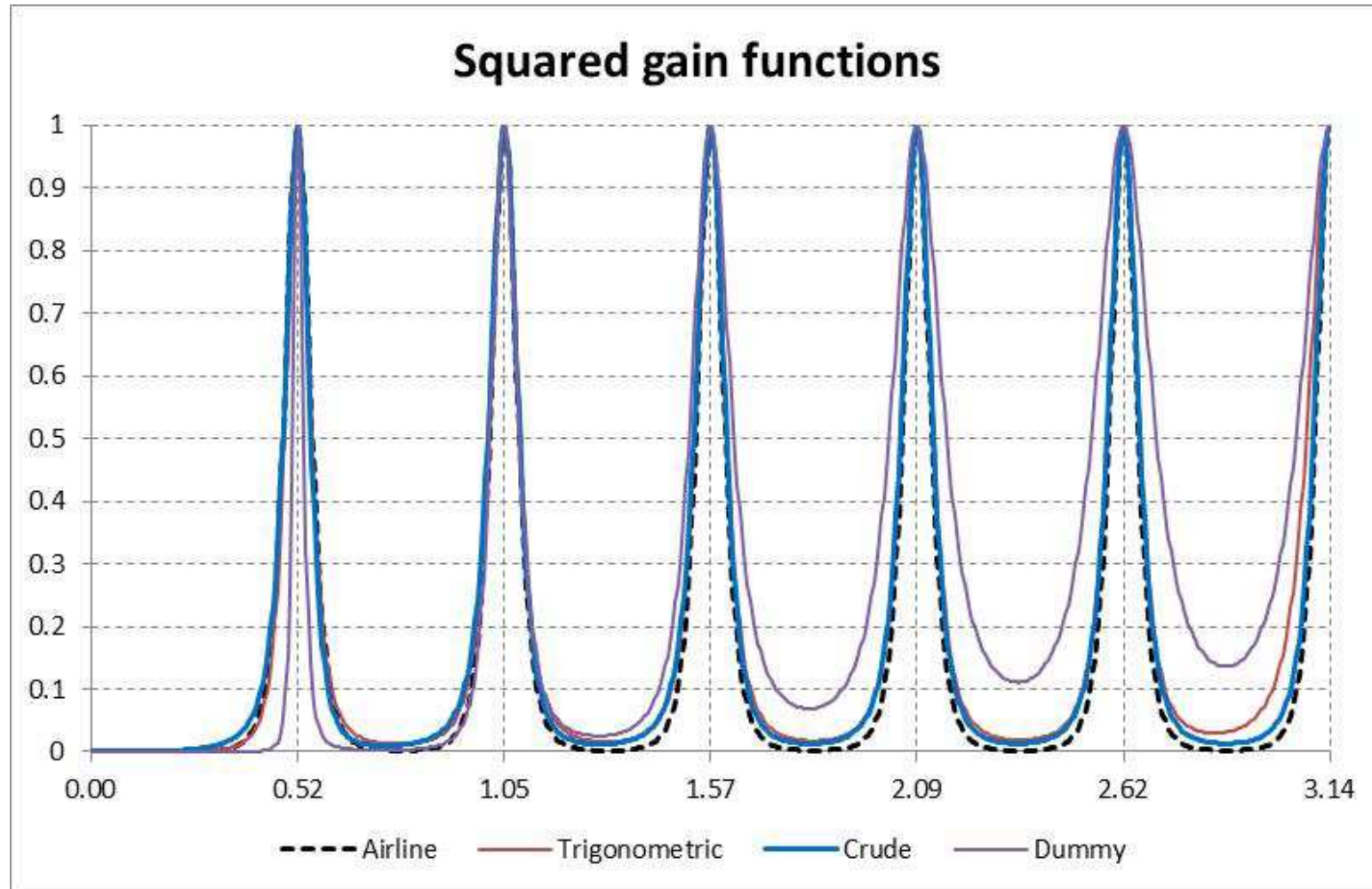
Reduced model

D: 1.00000 - B - B¹² + B¹³
MA: 1.00000 - 0.731773 B + 0.00617887 B² + 0.00561459 B³ + 0.00482616 B⁴ + 0.00377561 B⁵ + 0.00240978 B⁶ + 0.000656604 B⁷ - 0.00157990 B⁸ - 0.00442638 B⁹ - 0.00804923 B¹⁰ - 0.0126666 B¹¹ - 0.461117 B¹² + 0.345014 B¹³

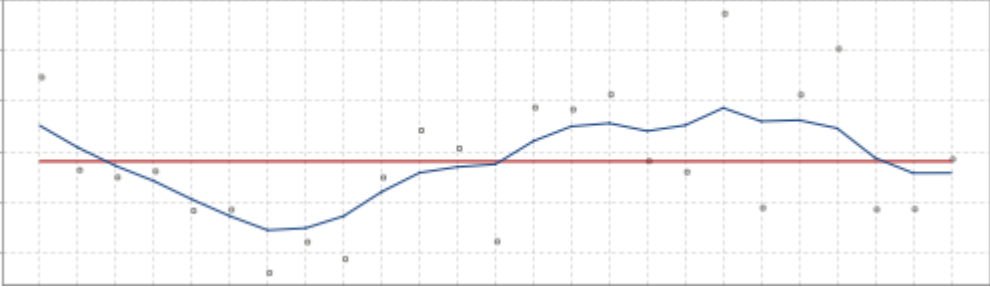
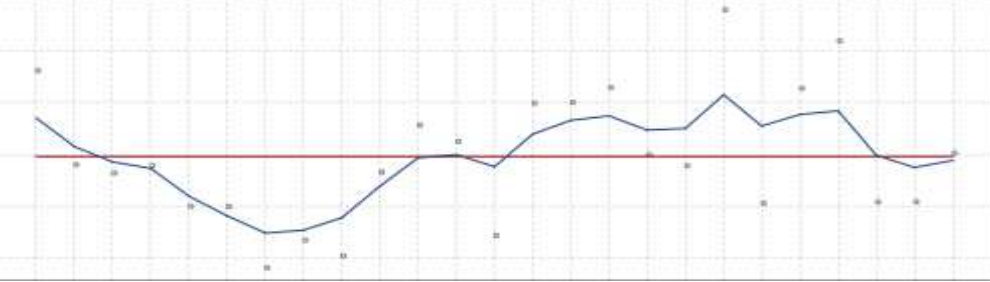
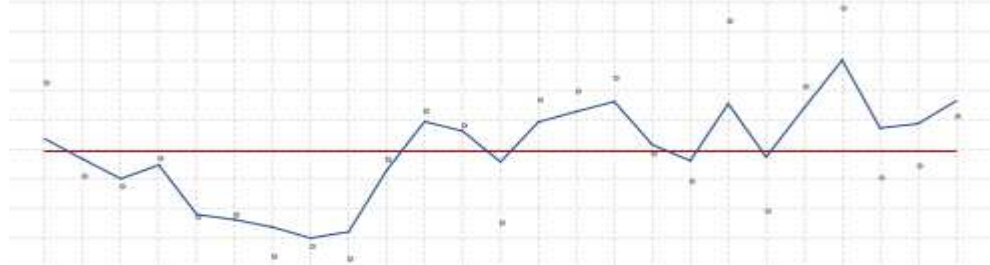


► Composante saisonnière

- fonctions de gain des filtres saisonniers (série INSEE 0016545073)



- Facteurs saisonniers (série INSEE 001654507, septembre)

Airline (AMB)	
Trigonometric (STS)	
Dummy (STS)	



► Comparaison STS-AMB. Conclusions

- Modèles réduits des STS \approx modèles airline
 - Exceptions:
 - ◆ Modèles non décomposables (AMB)
 - ◆ Contraintes sur les paramètres (variances = 0...)
 - ◆ STS plus complexes (cycles...)
- Décomposition canonique (AMB) \rightarrow
 - trend, composante saisonnière + stables
 - Irrégulier, séries désaisonnalisée + irrégulières
 - Approche + « mathématique » de la décomposition



3 Modèles alternatifs: exemples

- ▶ Objectif: élargir le champs (limité) des modèles habituels (→ options avancées du X11)
 - hétéroscédasticité des séries en fonction de la saison
 - Filtres différents suivant les saisons
 - Filtres différents suivant les fréquences
- ▶ Solutions
 - « Mixed airline model »: airline + dummies saisonnières spécifiques (1)
 - Généralisation du modèle de Harrison-Stevens (2)
 - Modèle « airline généralisé » (3)
- ▶ Algorithmes
 - WK standard non adapté aux cas 1 et 2. KF nécessaire.



3.1 Mixed airline model

► Forme espace-état

$$y_t = [Z_{arima} \quad Z_{nt}] \begin{bmatrix} \alpha_{arima,t} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{arima,t+1} \\ \alpha_{n,t+1} \end{bmatrix} = \begin{bmatrix} T_{arima} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{arima,t} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{arima,t} \\ \varepsilon_t \end{bmatrix}$$

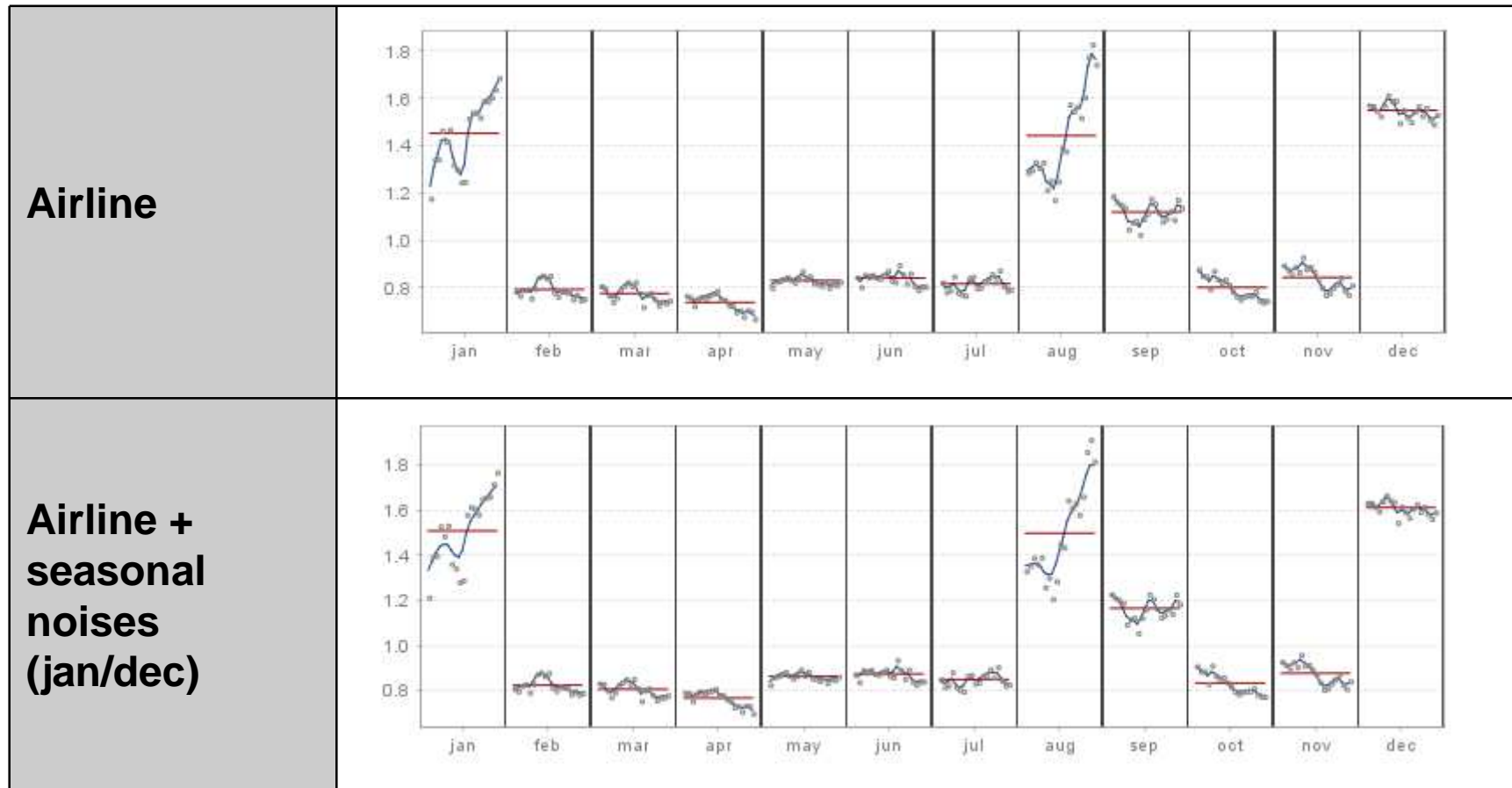
$$\begin{bmatrix} \mu_{arima,t} \\ \varepsilon_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Omega_{arima} & 0 \\ 0 & v_n \end{bmatrix}\right)$$

$$Z_{nt} = 1 \text{ si } t \bmod s \in S, 0 \text{ sinon}$$

(S est l'ensemble des saisons « perturbées »)



Mixed airline model (exemple)



	Janvier	Août
Airline		
Airline + seasonal noises (jan/août)		

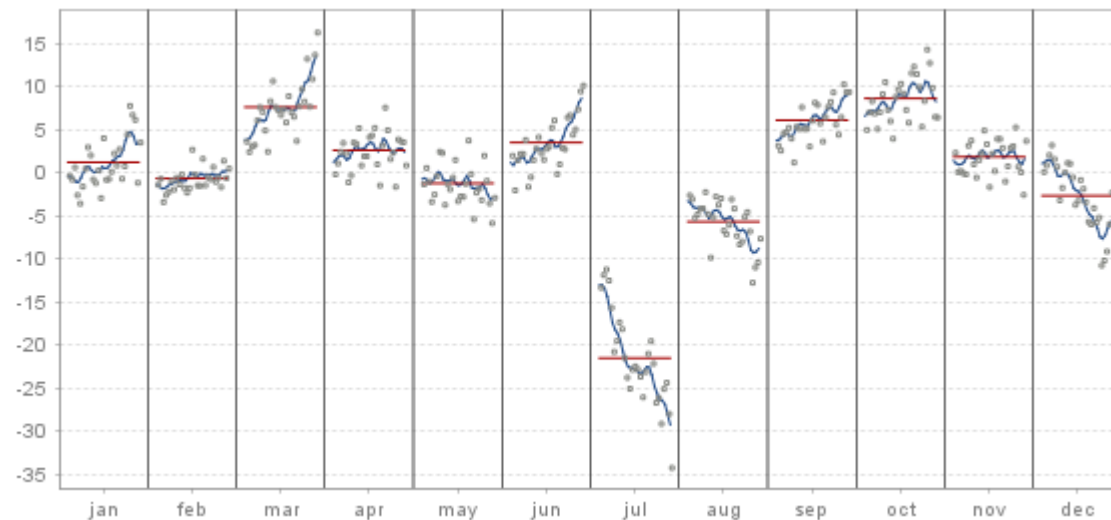


3.2 Généralisation du modèle Harrison-Stevens

$$\gamma_t = X_t \delta_t, \quad X_t \text{ dummies saisonnières,}$$
$$\delta_{i,t+1} = \delta_{i,t} + \varepsilon_{i,t}, \quad \begin{cases} \varepsilon_{i,t} \sim N(0, v_0), i \in S \\ \varepsilon_{i,t} \sim N(0, v_1), i \notin S \end{cases}$$

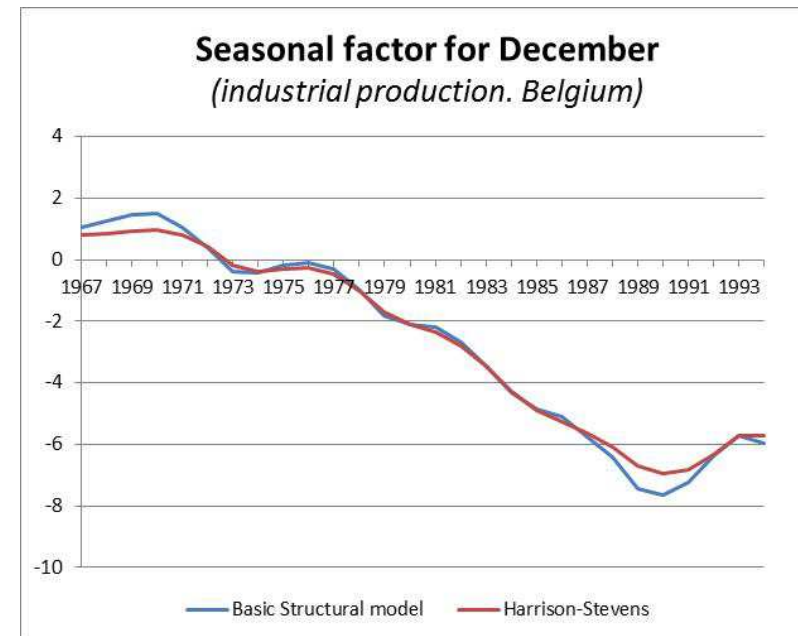
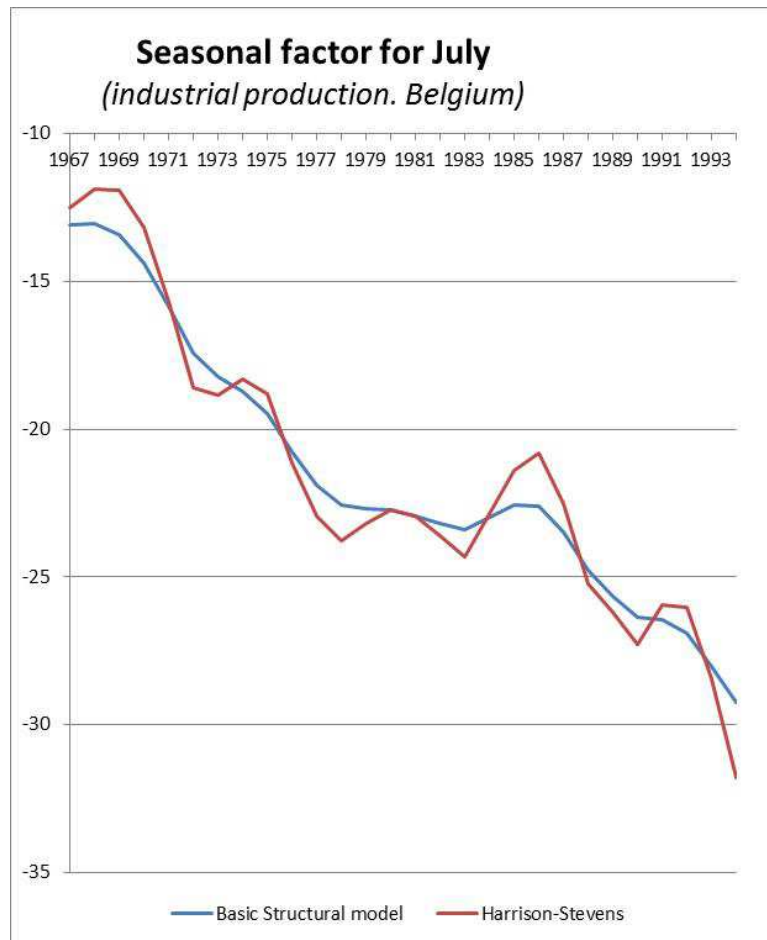
$v=0.0214$

Loglikelihood=-856.66



$v_0=0.0105$, $v_1=0.1094$,
Loglikelihood=-850.76

$S=\{\text{July}\}$



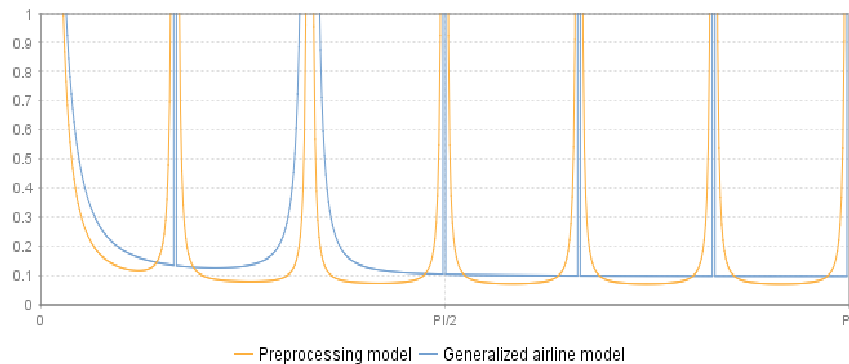
3.3 Airline généralisé

► **Modèle:** $1 - \theta B^s = (1 - \theta^{1/s} B) (1 + \theta^{1/s} B) \prod_{i=1}^5 (1 - 2\theta^{1/s} \cos \frac{2\pi i}{12} B + \theta^{2/s} B^2)$

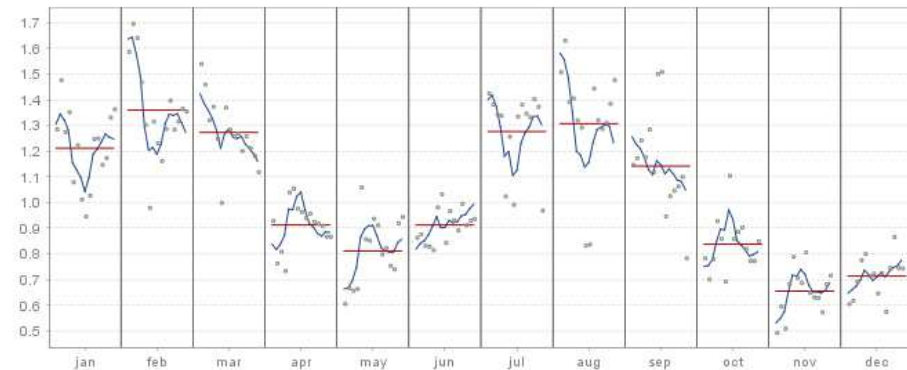


$= (1 - c_0 B)(1 + c_6 B) \prod_{i=1}^5 (1 - 2c_i \cos \frac{2\pi i}{12} B + c_i B^2)$

Spectres des modèles
airline et airline généralisé



Importation de chaussures. S-I ratio



4. Analyse des révisions

▶ Cadre:

- Décomposition canonique (SEATS)
- Révisions uniquement engendrées par les filtres de la décomposition
 - Modèle de régression fixé
 - Paramètres des modèles non ré-estimés

▶ Dérivation théorique

- WK / KF

▶ Exemples concrets



4.1 Analyse des révisions par WK

- ▶ Rôle des PsiE-weights

$$\hat{S}_t = \sum_{i=0}^{\infty} \xi_i^- \varepsilon_{t-i} + \sum_{i=1}^{\infty} \xi_i^+ \varepsilon_{t+i}$$

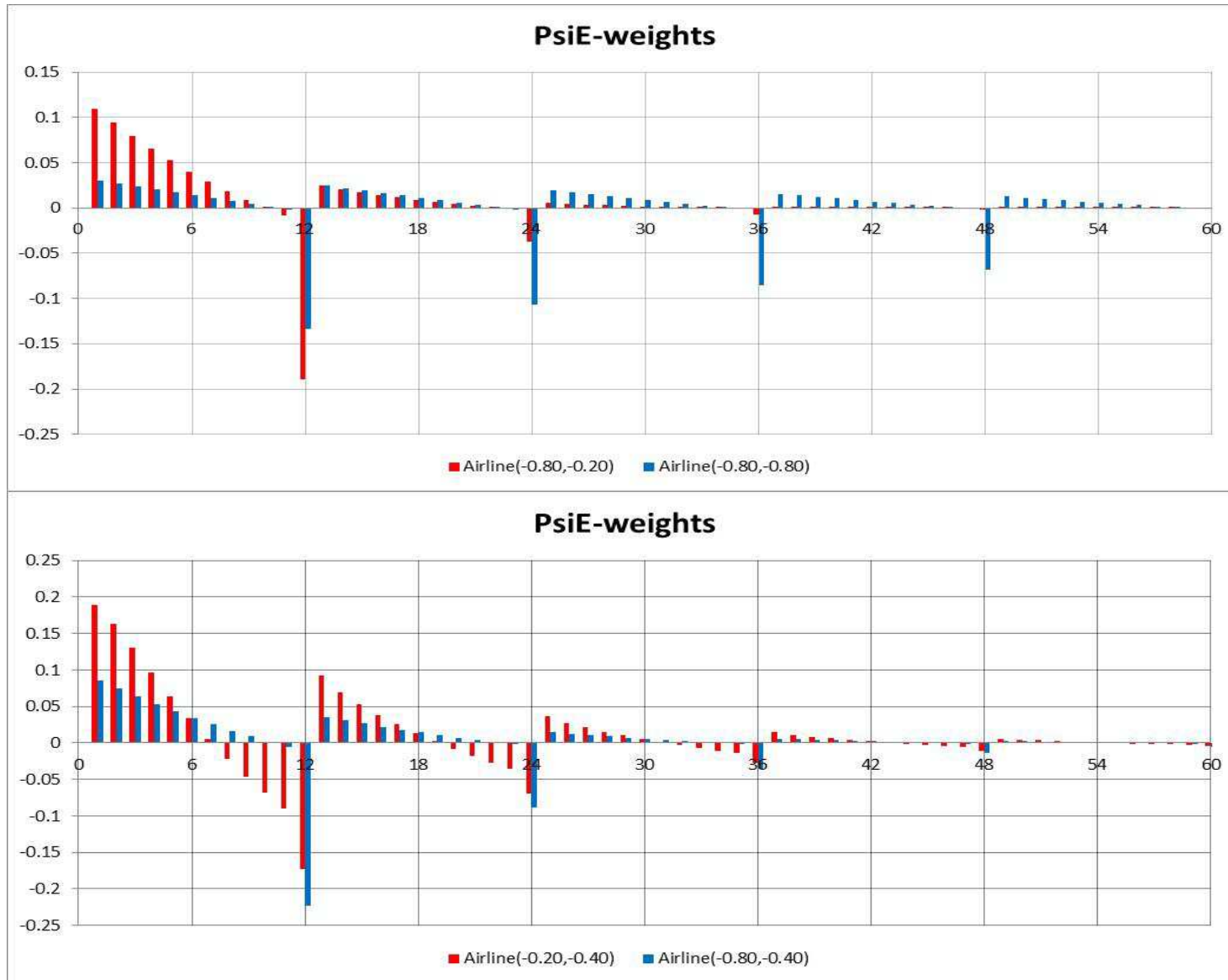
Révisions

- ▶ Révision du signal après k périodes:

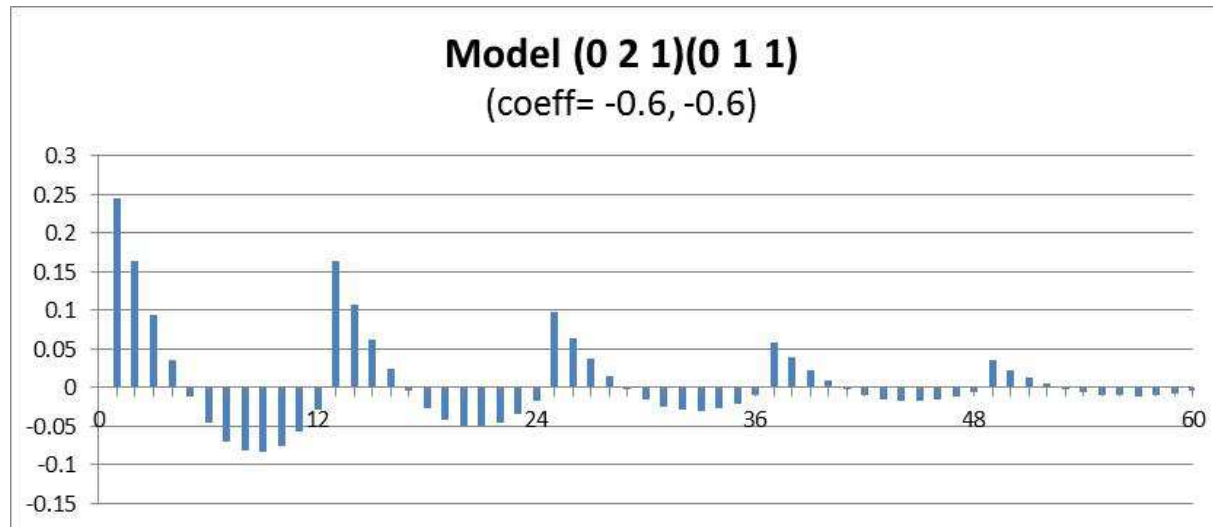
$$r(t|t+k) = \sum_{i=1}^k \xi_i^+ \varepsilon_{t+i}$$



► PsiE-Weights: modèles airline



▶ Autre modèle



- ▶ → ampleur et timing des révisions prévisibles
- ▶ Révisions explicables



4.2 Analyse des révisions par KF

- ▶ Modèle espace-état de la décomposition (rappel)

$$y_t = [Z_s \quad Z_n] \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{st+1} \\ \alpha_{nt+1} \end{bmatrix} = \begin{bmatrix} T_s & 0 \\ 0 & T_n \end{bmatrix} \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix}$$

- ▶ Révision en (t+k) du signal en t:

$$r(t|t+k) = \sum_{i=1}^k Z_s [E(\alpha_{st}|t+i) - E(\alpha_{st}|t+i-1)]$$



► Modèle « augmenté » (= fixed point smoothing)

$$y_{t+k} = [Z_s \quad Z_n \quad 0] \begin{bmatrix} \alpha_{st+k} \\ \alpha_{nt+k} \\ \alpha_{st} \end{bmatrix}$$

$$\alpha_{t+k+1} = \begin{bmatrix} \alpha_{st+k+1} \\ \alpha_{nt+k+1} \\ \alpha_{st} \end{bmatrix} = \begin{bmatrix} T_s & 0 & 0 \\ 0 & T_n & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \alpha_{st+k} \\ \alpha_{nt+k} \\ \alpha_{st} \end{bmatrix} + \begin{bmatrix} \mu_{st+k} \\ \mu_{nt+k} \\ 0 \end{bmatrix}$$

$$E(\alpha_{st}|t+k) = E(\alpha_{st}|t+k-1) + [0 \quad 0 \quad I]K_{t+k}v_{t+k}$$

v_{t+k} = innovation pour la période t+k

K_{t+k} = gain du filtre de Kalman du modèle augmenté

$$\gamma_{t+k} = [0 \quad 0 \quad Z_{s,t}]K_{t+k}$$

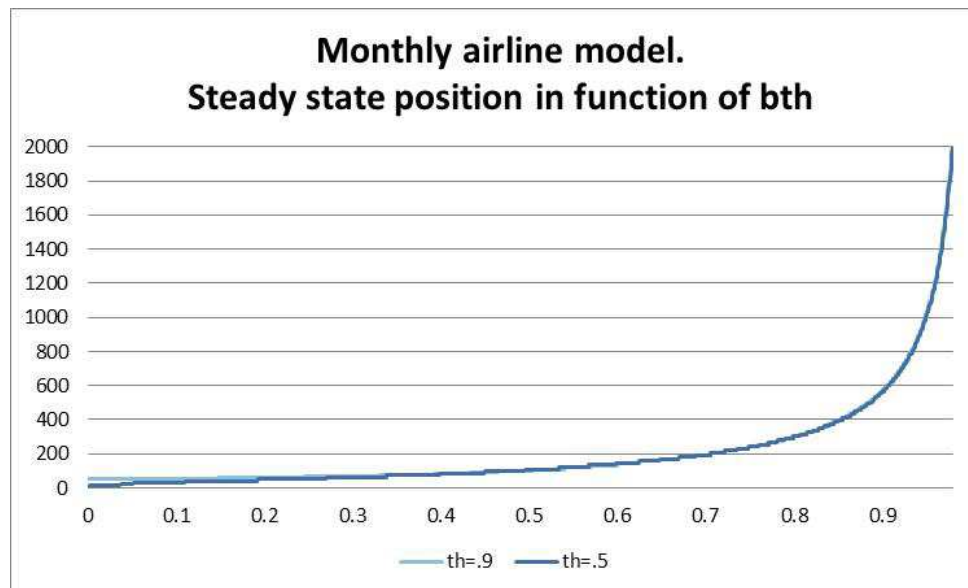
$$r(t|t+k) = \sum_{i=1}^k \gamma_{t+i} v_{t+i}$$

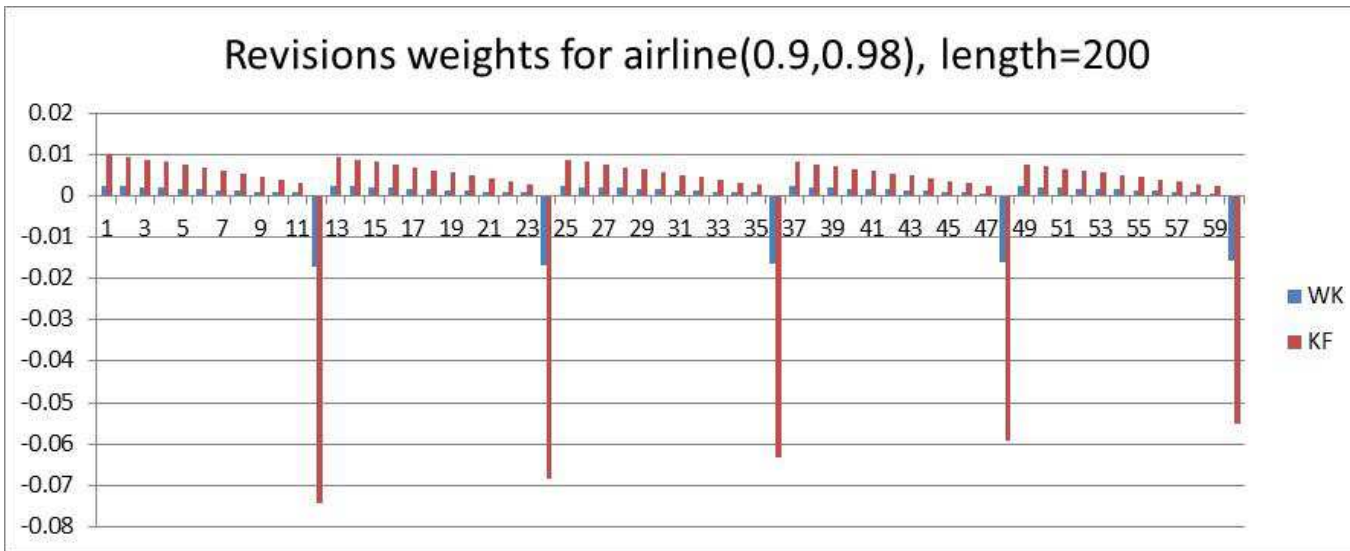
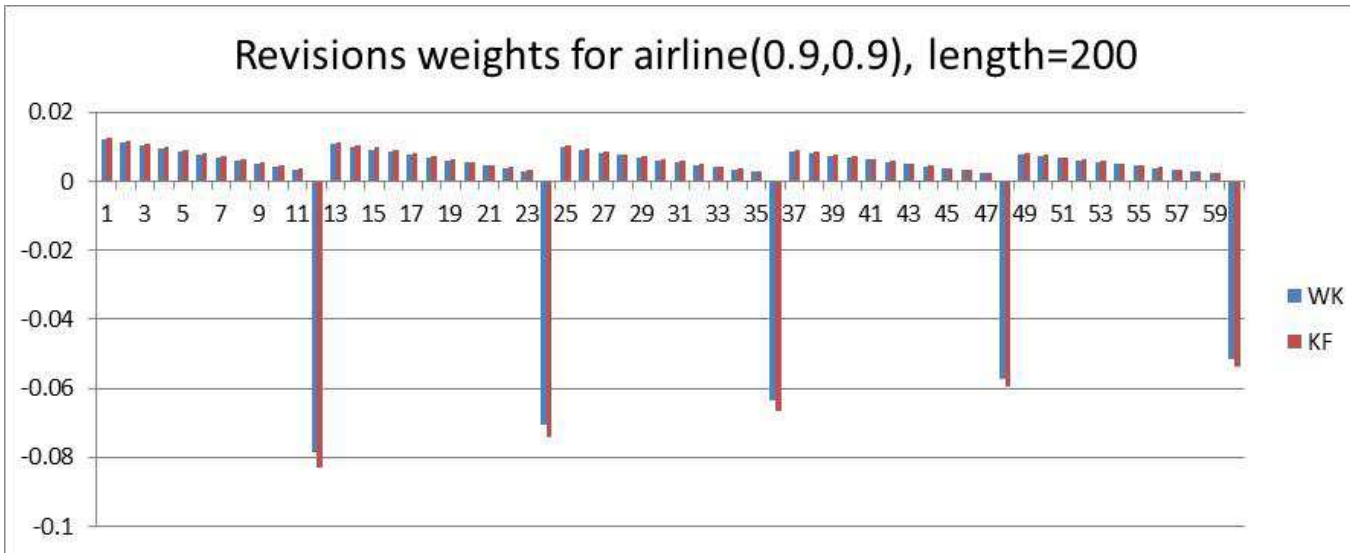


4.3 Comparaison des approches WK et KF

▶ $i > 0 : \xi_i^+ \approx \gamma_{t+i}$ (WK \approx KF)

le modèle est en état de (quasi-) équilibre à partir de la période t

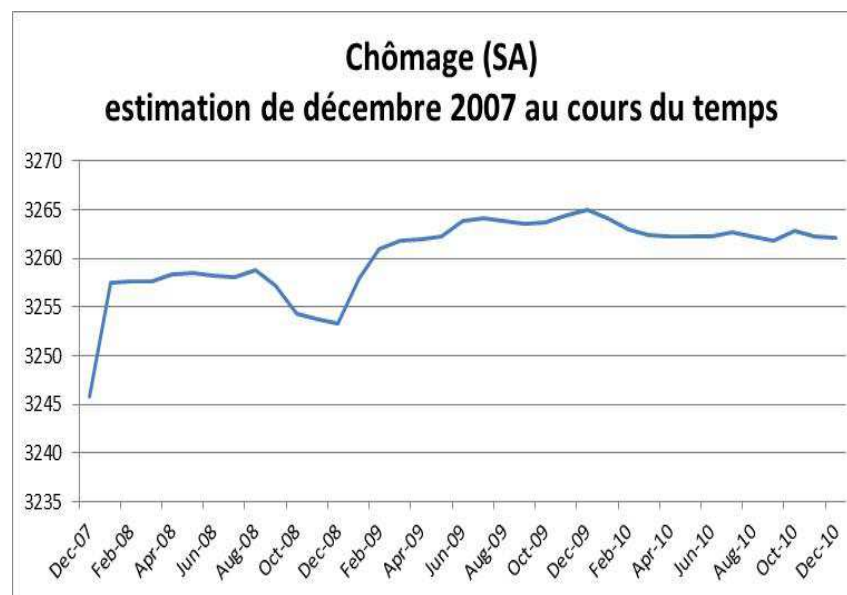




4.4 Analyse des révisions. Exemples

► Chômage en décembre 2007 (série INSEE :001572440)

Période	Innovations	PsiE-weights(WK)	Révisions	SA (décembre 2007)
Dec-07				3245.8
Jan-08	61.8	0.188	11.6	3257.4
Feb-08	2.0	0.125	0.2	3257.7
Mar-08	-1.7	0.071	-0.1	3257.6
Apr-08	31.3	0.026	0.8	3258.4
May-08	-15.6	-0.011	0.2	3258.5
Jun-08	8.9	-0.038	-0.3	3258.2
Jul-08	3.3	-0.056	-0.2	3258.0
Aug-08	-12.6	-0.066	0.8	3258.8
Sep-08	25.5	-0.067	-1.7	3257.1
Oct-08	48.3	-0.059	-2.9	3254.3
Nov-08	10.5	-0.044	-0.5	3253.8
Dec-08	27.8	-0.020	-0.6	3253.2
Jan-09	33.7	0.141	4.7	3258.0
Feb-09	31.4	0.093	2.9	3260.9
Mar-09	17.8	0.053	0.9	3261.8
Apr-09	9.3	0.020	0.2	3262.0
May-09	-36.6	-0.006	0.2	3262.3
Jun-09	-58.3	-0.026	1.5	3263.8
Jul-09	-9.2	-0.039	0.4	3264.1
Aug-09	6.9	-0.046	-0.3	3263.8
Sep-09	6.3	-0.046	-0.3	3263.5
Oct-09	-4.6	-0.041	0.2	3263.7
Nov-09	-21.2	-0.031	0.6	3264.4
Dec-09	-42.2	-0.014	0.6	3265.0



► Prod. ind. en décembre 2013 (série INSEE: 001654235)

	Innovation	PsiE-weight (WK)	Impact sur déc. 2013 (SA)
01/01/2014	-1.2%	0.092	-0.11%
01/02/2014	0.3%	0.078	0.02%
01/03/2014	-0.3%	0.062	-0.02%
01/04/2014	1.6%	0.047	0.08%
01/05/2014	-3.4%	0.031	-0.11%
01/06/2014	1.9%	0.016	0.03%
01/07/2014	0.7%	0.001	0.00%
01/08/2014	-0.2%	-0.013	0.00%
01/09/2014	0.7%	-0.027	-0.02%
01/10/2014	-1.1%	-0.040	0.04%
01/11/2014	-1.6%	-0.053	0.08%
01/12/2014	3.6%	-0.130	-0.47%



- Modèle Airline
(→Révision après 1 an)
- Interprétation « intuitive »

Forte « surprise » positive en décembre 2014

Facteur saisonnier de décembre à revoir à la hausse (y compris en 2013, 2012 ...)

Révision à la baisse de décembre 2013



4.5 Analyse de révisions: conclusions

- ▶ L'approche par modélisation permet de comprendre et d'expliquer les révisions des séries désaisonnalisées
- ▶ L'approche semi-infinie de WK (SEATS) fournit une information presque exacte sauf dans le cas de modèles quasi-non inversibles; dans de tels cas, il faut recourir à KF.



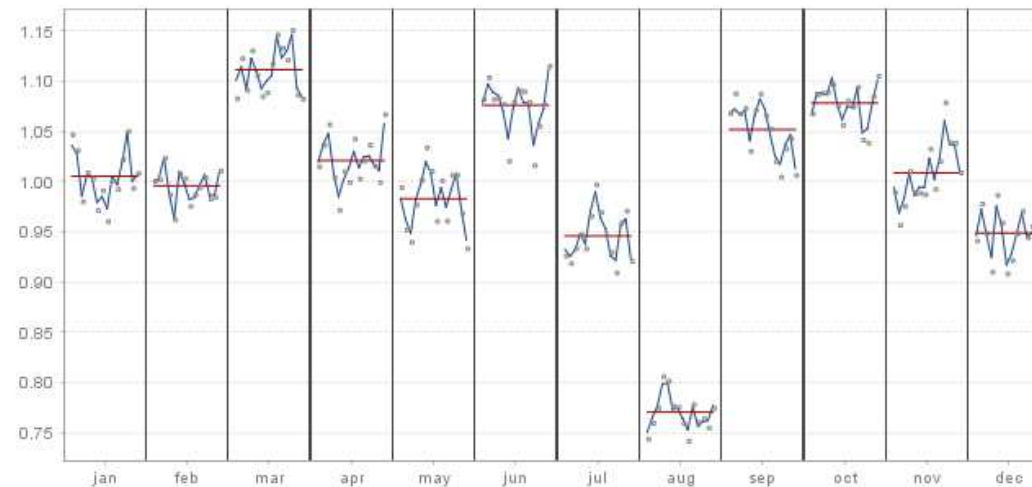
5. Limites de la modélisation

- ▶ AMB → Composante saisonnière parfois erratique
 - Exportations de la Belgique vers la France

Arima model

[(1,1,0)(1,1,1)]

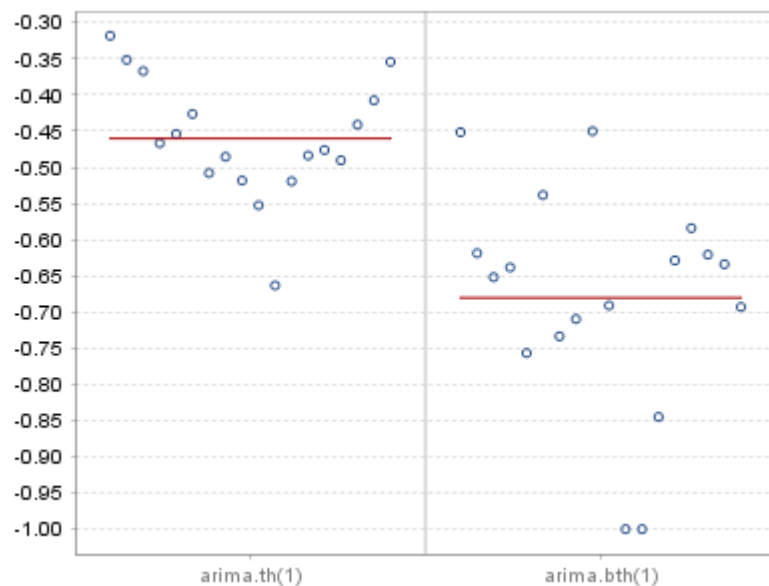
	Coefficients	T-Stat	P[T > t]
Phi(1)	0.6382	9.98	0.0000
BPhi(1)	-0.2866	-2.89	0.0045
BTheta(1)	-0.8734	-13.53	0.0000



► Estimations parfois instables (AMB, STS...)

- Production industrielle (séries INSEE 001654235)

Modèles estimés sur
des périodes glissantes
(8 ans)



Modèle estimé sur
l'ensemble de la période
(25 ans)

Arima model
[(0,1,1)(0,1,1)]

	Coefficients	T-Stat	P[T > t]
Theta(1)	-0.4870	-9.22	0.0000
BTheta(1)	-0.6769	-14.23	0.0000



► Multiples solutions locales pour un même modèle (modèles complexes)

TRAMO

Loglikelihood = -692.0892
 Standard error of the regression (ML estimate) = 2.6573
 AIC = 1412.1783
 AICC = 1413.7168
 BIC = 1463.4598
 BIC (Tramo definition) = 2.2102
 Hannan-Quinn = 1432.7289

ARIMA model [(3,1,1)(0,1,1)]

Parameter	Value	Std error	T-Stat	P-value
Phi(1)	1.4458	0.1361	10.62	0.0000
Phi(2)	0.7085	0.1079	6.56	0.0000
Phi(3)	0.2595	0.0618	4.20	0.0000
Th(1)	0.9852	0.1309	7.53	0.0000
BTh(1)	-0.5013	0.0575	-8.72	0.0000

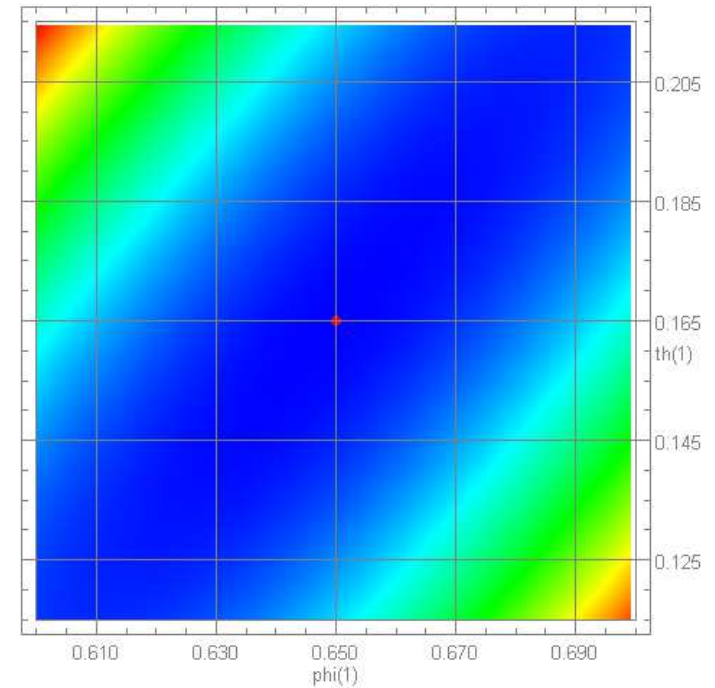
X12

Loglikelihood = -692.9545
 Standard error of the regression (ML estimate) = 2.66634
 AIC = 1413.9089
 AICC = 1415.4474
 BIC = 1465.1904
 BIC (Tramo definition) = 2.2170
 Hannan-Quinn = 1434.4594

ARIMA model [(3,1,1)(0,1,1)]

Parameter	Value	Std error	T-Stat	P-value
Phi(1)	0.6417	0.7752	0.83	0.4085
Phi(2)	0.3341	0.3730	0.90	0.3712
Phi(3)	0.0800	0.1898	0.42	0.6736
Th(1)	0.1562	0.7751	0.20	0.8404
BTh(1)	-0.4986	0.0520	-9.58	0.0000

Fonction de vraisemblance



Conclusions

- ▶ Richesse de l'approche de la désaisonnalisation par modélisation
 - Solutions variées, flexibles
 - Cadre d'analyse riche
 - Choix personnel:
 - Décomposition AMB + filtres WK pour 95% des séries
 - Modèles structurels + filtres KF pour se faire plaisir
- ▶ Limites de la modélisation
 - Nécessité de modèles robustes (= parcimonieux)
 - Ré-estimation mesurée des modèles
 - [Utilisation non-paramétrique des modèles]
- ▶ JD+ outil varié et puissant

