Exploitation des déclarations douanières pour un indice de prix

Journées suisses de la statistique, 20.11.2017

Christophe Matthey, Office fédéral de la statistique

Sommaire

1. Objectifs
2. Représentation standard vs déclarations
3. Extraction d'information
4. Séries de prix
5. Conclusions, futurs travaux

Objectifs

- Objectifs :

1. Construction de séries de prix basées sur les déclarations douanières (données DGD - OFS)
2. Evaluation des séries obtenues
3. Construction d'un indice de prix

- Informations disponibles : valeur exportée (CHF), quantité exportée (kg), pays de destination, description de la marchandise exportée
- Prix : CHF / kg par mois/trimestre (valeur moyenne)

Objectifs

- Définition séries de prix : 1 médicament x 1 pays de destination (prix différenciés, négociation avec états)

Questions :

⇒ **Quelle info sur médicaments au sein des descriptions, comment l'extraire ?**

⇒ **Comment évaluer la qualité des séries obtenues ?**

⇒ **Comment combiner les séries au sein d'un indice ?**

Représentation standard vs réalité

- Représentations standards recherchées :

Type	Nom médicament	Dosage p.a.*	Unité p.a.
Propriétaire	Dafalgan	500	mg
Générique	Paracetamol Mepha	500	mg

Forme	Représentation(s) principale(s)	Alternatives
1. Tablettes, suppositoires, poudres (FDA: 64%)	MG	MEQ, IU
	Si plusieurs molécules : X1 / X2…/ Xn MG ou Somme(Xi) MG	
2. Perfusions/injections (21%)	MG/ML, MG/ ampoule, %	MEQ/ML, MCI/ML, IU/ML, KBQ/ML, MMOL/ML
3. Autres liquides, ex. sirop, gouttes (8%)	MG/ML, %	
4. Crème, lotion (3%)	%	MG/G, MG
5. Patch dermique (1%)	MG/H	
6. Aerosol, spray (1%)	MG/inhalation	%, MG/ML

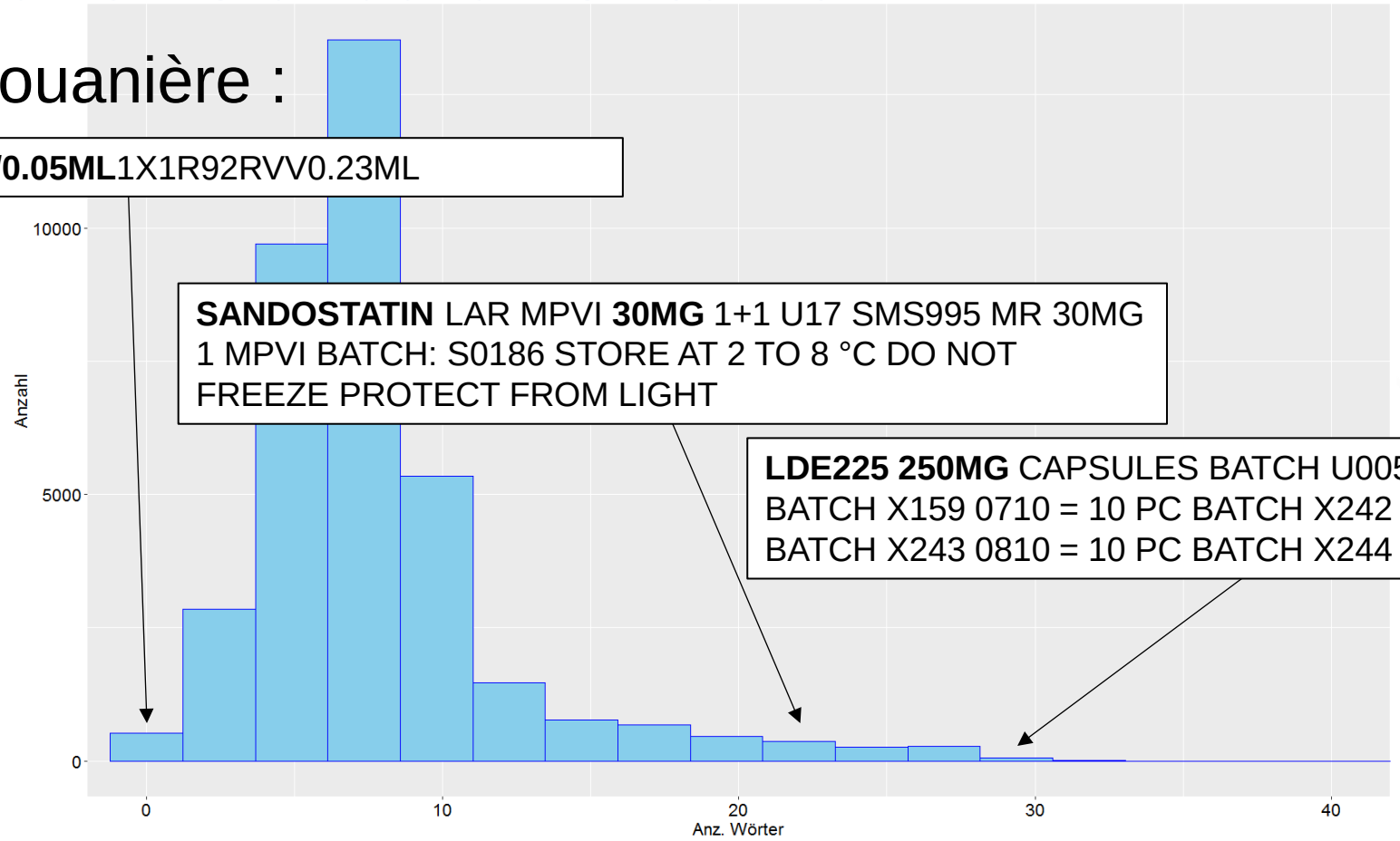
Représentation standard vs réalité

- Réalité douanière :

LUCENTIS LIVI**0.5MG/0.05ML**1X1R92RVV0.23ML

SANDOSTATIN LAR MPVI **30MG** 1+1 U17 SMS995 MR 30MG
1 MPVI BATCH: S0186 STORE AT 2 TO 8 °C DO NOT
FREEZE PROTECT FROM LIGHT

LDE225 250MG CAPSULES BATCH U005 0710 = 10 PC
BATCH X159 0710 = 10 PC BATCH X242 0810 = 10 PC
BATCH X243 0810 = 10 PC BATCH X244



Extraction info sur médicament

1. Directe : exploitation de régularités observées dans les descriptions, regex ad hoc

   => Elément clé : manière de remplir les déclarations stable

2. Basée sur une liste: Drugbank (université de l'Alberta), info sur molécule (nom générique), produits et dosages concernés + infos sur ATC,

   => Elément clé : actualité/exhaustivité Drugbank

Extraction info sur médicament

3. Combinée : p.ex. sélection ssi même info (molécule et/ou dosage), extraction directe utilisée pour choisir entre plusieurs infos selon DB

=> Couverture variable selon méthode d'extraction (2013-2014)

Info extraite	Nb		Valeur totale		Quantité totale	
WARENBEZEICHNUNG	67'406	100.0%	111'178'712'774	100.0%	314'979'947	100.0%
WARENBEZEICHNUNG hors RESEARCH…	60'220	89.3%	110'739'855'051	99.6%	314'640'037	99.9%
Extraction directe	15'373	22.8%	106'391'021'027	95.7%	304'179'950	96.6%
Extraction Drugbank	5'150	7.6%	88'853'708'066	79.9%	126'674'135	40.2%
Extraction combinée	4'260	6.3%	87'760'753'200	78.9%	119'912'003	38.1%

Séries de prix

1. Nombre variable selon méthode d'extraction (2013-2014)

Nb séries potentielles	Nb
Extraction directe	51'961
Extraction Drugbank	22'980
Extraction combinée	20'898

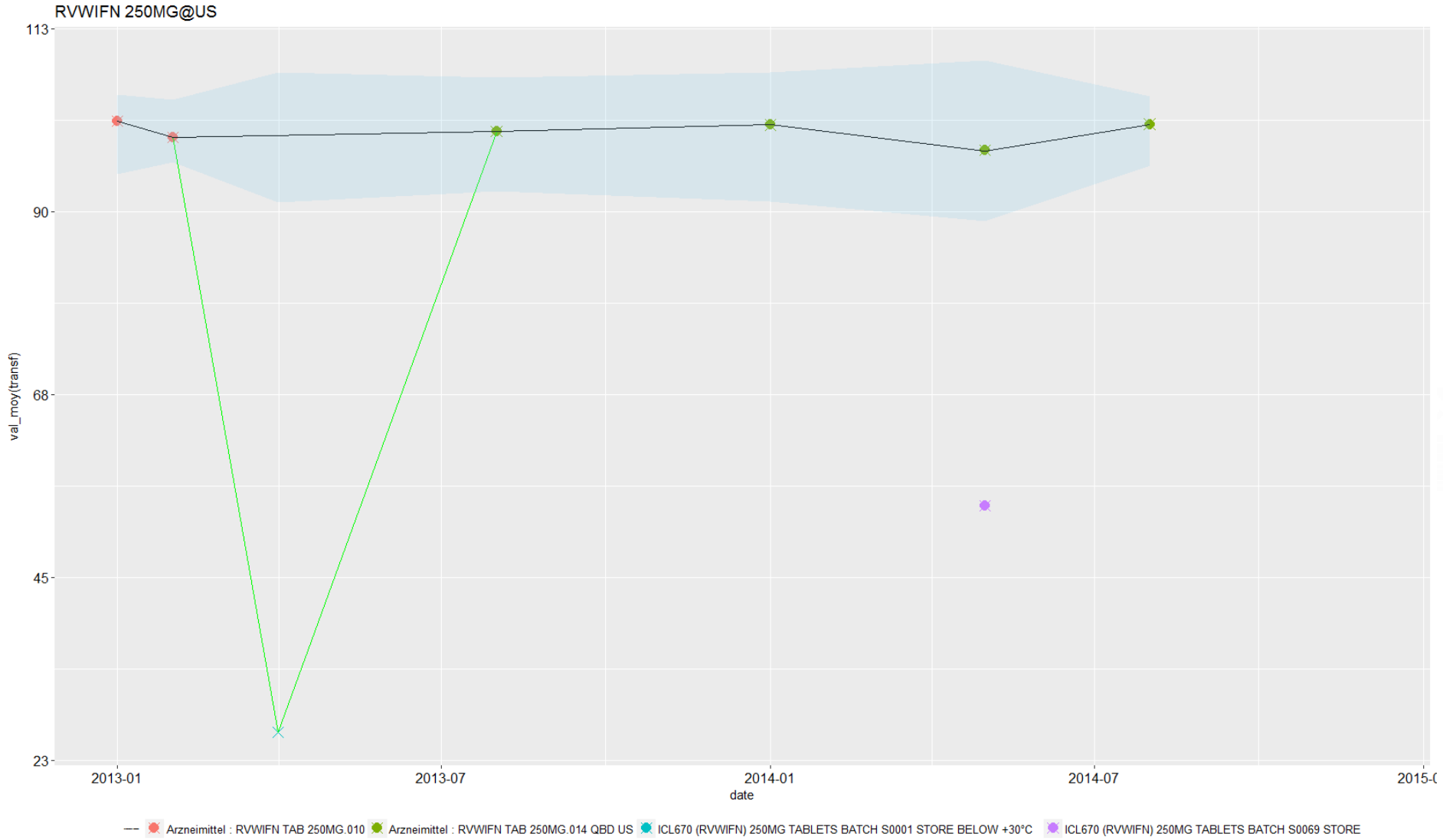
2. Qualité à évaluer, observations/séries problématiques à éliminer

- filtres au niveau des descriptions / séries complètes, prise en compte de la dimension temporelle,
- descriptions : écart par rapport à MAD (fenêtre glissante),
- série : écart maxmin, nb de descriptions, nb d'apparitions,…

Séries de prix

Quelques problèmes rencontrés

- Outliers (cf slide suivante)
- Séries «hub»
- Séries multi-produits


RVWIFN 250MG@US
113
90
68
45
23
val_moy(transf)
2013-01
2013-07
2014-01
2014-07
2015-(
date
Arzneimittel : RVWIFN TAB 250MG.010
Arzneimittel : RVWIFN TAB 250MG.014 QBD US
ICL670 (RVWIFN) 250MG TABLETS BATCH S0001 STORE BELOW +30°C
ICL670 (RVWIFN) 250MG TABLETS BATCH S0069 STORE

Conclusions intermédiaires, futurs travaux

- 1ers résultats prometteurs, gains attendus dans la couverture des produits/pays d'exportation
- MAIS forte dépendance envers la manière de remplir les déclarations (peut varier d'un exportateur à l'autre) et peu de moyens de l'influencer
- Tests à poursuivre sur les données mensuelles 2016 – 2017
  - Développement d'un indice
  - Evaluation des filtres (impact sur indice ?)
- Réflexions/tests pour la mise en production (2018 ?)

Questions