Etude de la dépendance entre la situation de tension du MA et la formation des prix du MA

# Table des matières

[Table des matières 1](#_Toc497754082)

[1. Plusieurs sous hypothèses 2](#_Toc497754083)

[1. Hypothèse H1 : à situation d’alerte différente les prix et marges sont comparables 2](#_Toc497754084)

[1. Alert vs PRE/PMP/PME et Alert vs Peak Daily Margin 2](#_Toc497754085)

[2. Downgraded vs PRE/PMP/PME et Downgraded vs Peak Daily Margin 2](#_Toc497754086)

[3. PP1/PP2 vs PRE/PMP/PME et PP1/PP2 vs Peak Daily Margin 2](#_Toc497754087)

[2. Hypothèse H2 : Les périodes de transitions avant une alerte présentent des prix homogènes avec les prix des périodes précédentes 2](#_Toc497754088)

[1. Trois zones : pré-alert, intra-alert et post-alert. Il s’agit de de faire une première critique de l’estimation du niveau d’alerte levée par RTE ou si des signes avant-coureurs peuvent être quantifiés pour anticiper de tels signaux. 2](#_Toc497754089)

[3. Hypothèse H3 : Les différents niveaux d’alertes sont indépendants et présentent des réponses de prix différentes 2](#_Toc497754090)

[2. Données utilisées 2](#_Toc497754091)

[1. Type de données 2](#_Toc497754092)

[1. Prix sur le mécanisme d’ajustement 2](#_Toc497754093)

[2. Signaux d’alertes 2](#_Toc497754094)

[2. Plage de données 3](#_Toc497754095)

[3. Procédure 4](#_Toc497754096)

[1. Description Statistique 4](#_Toc497754097)

[2. Protocole 4](#_Toc497754098)

[1. Classification 4](#_Toc497754099)

[2. Les tests 4](#_Toc497754100)

[4. Applications aux données 5](#_Toc497754101)

[1. Description Statistique 5](#_Toc497754102)

[1. Courbe de charge RTE 5](#_Toc497754103)

[2. Résultats Détaillés 6](#_Toc497754104)

[1. Classifications Utilisées 6](#_Toc497754105)

[2. Test de normalité 7](#_Toc497754106)

[3. Analyse des impacts de la situation d’alerte sur les prix du Mécanisme d’ajustement 9](#_Toc497754107)

[5. Conclusion 14](#_Toc497754108)

[6. Références 14](#_Toc497754109)

# Plusieurs sous hypothèses

## Hypothèse H1 : à situation d’alerte différente les prix et marges sont comparables

### Alert vs PRE/PMP/PME et Alert vs Peak Daily Margin

### Downgraded vs PRE/PMP/PME et Downgraded vs Peak Daily Margin

### PP1/PP2 vs PRE/PMP/PME et PP1/PP2 vs Peak Daily Margin

## Hypothèse H2 : Les périodes de transitions avant une alerte présentent des prix homogènes avec les prix des périodes précédentes

### Trois zones : pré-alert, intra-alert et post-alert. Il s’agit de de faire une première critique de l’estimation du niveau d’alerte levée par RTE ou si des signes avant-coureurs peuvent être quantifiés pour anticiper de tels signaux.

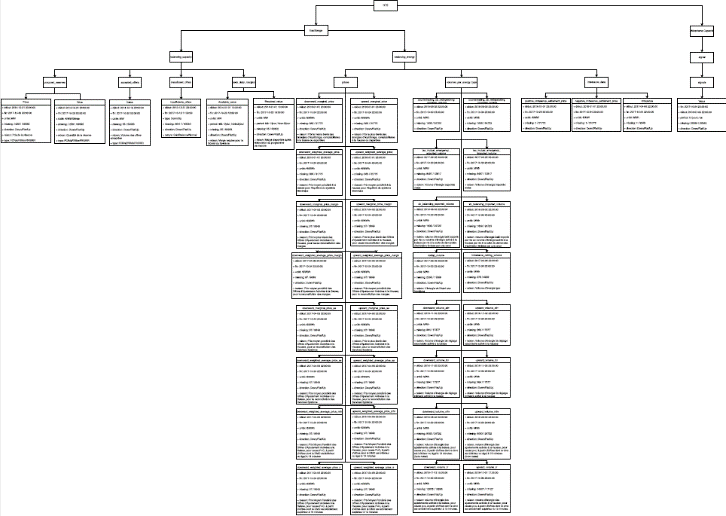
* + PRE/PMP/PMEpré-alert vs PRE/PMP/PMEintra-alert
  + PRE/PMP/PMEintra-alert vs PRE/PMP/PMEpost-alert

## Hypothèse H3 : Les différents niveaux d’alertes sont indépendants et présentent des réponses de prix différentes

* + PP1/PP2 vs Alert/Downgraded

# Données utilisées

## Type de données

Une cartographie expliquant la structure des données fournies par RTE est disponible [ici](Annexes/cartographie_données_dépendance.pdf). Les définitions de chaque sous ressource est disponible sous le label « raison ». Pour avoir une arborescence plus complète des ressources utilisées il est possible de se référer à la [page github suivante](https://github.com/mathiaHa/MA-Analysis/tree/rteb/RTE).

## Plage de données



Les données utilisées sont issues entièrement de [RTE](https://data.rte-france.com/home) des apis suivantes :

* Equilibrage/balancing\_capacity
* Equilibrage/balancing\_energy

La colonne « manquant » indique le nombre de pas demi-horaire manquant sur le nombre total de pas demi-horaire entre la valeur début et fin. Ce nombre est à prendre avec précautions. Cela englobe à la fois une donnée manquante dû à un bug lors de la récupération des données, un bug de la part de RTE, une absence de relève, des stockages différents pas au pas demi-horaire ou simplement l’absence d’acceptation d’offre par RTE pour des offres acceptées. Néanmoins il peut donner une idée sur le nombre de pas demi horaire ou des offres sont activées.

# Procédure

## Description Statistique

Objectif : Les tests statistiques utilisés dans l’étude des liaisons entre variables reposent sur des hypothèses auxquelles il est possible de répondre par une description statistique des données.

Nous fournirons un tableau récapitulatif des statistiques de bases pour chaque données continues ainsi que l’ensemble des métadonnées pour chaque intervalle de temps.

## Protocole

Il est très rapidement apparu qu’il serait difficile d’obtenir des résultats similaires en fonction des ressources et sous ressources utilisées. Nous allons donc dans un premier temps générer un grand nombre de possibilité de classification pour chaque sous ressource puis trier en fonction de la méthode de test et agréger en sélectionnant le résultat qui se répété le plus souvent par sous ressource et par type de test.

Pour répondre à l’hypothèse nulle :

**La distribution d’une sous ressource est la même quel que soit le niveau de tension du système électrique**.

### Classification

En affichant les distributions de sous ressource selon différentes classification il est clair que l’on peut atteindre une sorte d’« over-fitting » qui pourraient fausser la classification. Nous choisirons de façon dynamique la meilleure classification (ou du moins celle qui donne le résultat le plus tranché) par sous-ressource. Nous pouvons retenir les classifications suivantes :

* ["plage","value\_type"]
* ["month","value\_type"]
* ["month","plage","value\_type"]
* ["month","plage","value\_type","bday"]
* ["month","plage","value\_type","direction"]
* ["month","plage","value\_type","direction","bday"]

### Les tests

Il est nécessaire pour utiliser des tests paramétriques de faire l’hypothèse de distributions normales. C’est pour cela que le premier test que nous ferons sera de vérifier la normalité de chaque sous-ressource en fonction de chaque classification.

Une fois ce test réalisé il est possible de sélectionner des tests paramétriques ou non-paramétriques pour tester l’appartenance à une même population de distribution sous différentes situations de tensions électriques. Nous verrons dans la partie résultats que l’étude mène à des distributions non-paramétriques.

Nous reproduirons donc pour chaque sous ressource, chaque classification et chaque niveau de tension du systèmes les tests sur les distributions. Dans un premier temps, la stratégie conservée est de sélectionner la classification présentant la plus haute fréquence de rejet ou d’acceptation de l’hypothèse nulle par méthode. Cette stratégie de non-sélection est aussi motivé par la nature variante des signaux enregistré par RTE. Certains signaux sont intrinsèquement enregistrés pour un système à la hausse. L’ajout ou le retrait de la classification « direction » ne change rien au résultat. Nous pouvons prendre l’exemple de « downward\_volume\_mfrr ». Ce non choix peut néanmoins mener à des résultats sans homogénéité si une classification différente est retenue par sous ressource. Mais cela peut aussi mettre en avant des comportements avec des périodicités différentes (si on estime le résultat pertinent).

# Applications aux données

Objectif : Répondre à la problématique principale en répondant successivement à toutes les hypothèses sous-jacentes. Pour cela nous allons les différents tests pour tous les cas présentés en parties 1.

## Description Statistique

### Courbe de charge RTE

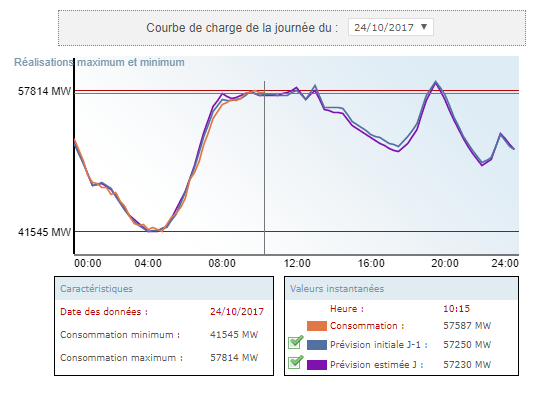
Cette étude est basée en grande partie sur la courbe de charge à la maille France agrégée et publiée par RTE. Cette courbe présente des pics de consommations à 8h-12h et 17h-20h qui donnent lieu à des dispositions de RTE avec les marges de production électrique (« peak\_daily\_margins »). Le marché de l’électricité est un marché sensible où beaucoup d’estimations sont faites jusqu’à J-1 et J. Le but est d’estimer au mieux la consommation pour être capable de la couvrir entièrement. Néanmoins il subsiste des écarts entre production et consommation dont une partie de la responsabilité incombe au « responsable d’équilibre » qui seront sanctionnés au « prix des écart » (« imbalance\_data »). Ces prix sont fixés en fonctions des prix des offres mises sur le Mécanisme d’Ajustement. En calculant les Prix Moyens Pondérés à la baisse et à la hausse (« prices »), on peut en déduire le précédent depuis le passage au « single price ». Quand l’équilibre du réseau est en danger, des procédures mises en place par RTE, signalent des mesures prises pour solutionnés ces événements par des alertes ou messages de mode dégradé (« insufficients\_offers »). Enfin, la France a un marché thermosensible qui en période de grand froid présente des journées de fortes tensions sur le marché de l’électricité, ces journées sont signalées « PP1/PP2 » (« signals »). Pour plus de détails il est possible de consulter la page [cartographie des données suivantes](Annexes/cartographie_données_dépendance.pdf).

Figure Courbe de Charge à la maille France

## Résultats Détaillés

Les résultats de ce test sont présentés dans le fichier [Excel Population\_Independance](Results/Population_Independance.xlsx). Les résultats sont répartis entre les 7 feuilles de classeur. Chaque feuille correspond à l’analyse de faite pour une classification et regroupe les résultats des tests de normalité, de Kruskal-Wallis et de Mann-Whithney.

Remarque : Tous les tests statistiques ont été mené avec un risque primaire de 5%. Il est important de garder à l’esprit que tous les résultats contiennent une part d’erreur d’autant au moment d’agréger les résultats après utilisations des classifications.

### Classifications Utilisées

La première recherche fut de déterminer s’il était possible d’isoler une classification pour obtenir des clusters de points présentant une distribution normale. Cette condition de normalité est une hypothèse de base des tests dits « paramétriques ». Reformuler la recherche est de diviser l’ensemble des observations afin de trouver des schémas qui se répètent et de pouvoir conclure à des comportements normaux sur des sous-ensembles et utilisé les tests paramétriques par sous-ensembles.

Le marché présente des spécificités telle que de forte périodicité présentée dans la partie sur la courbe de charge à la maille France. De plus, il n’est pas possible de modifier des valeurs d’offres déposées à l’intérieur des plages horaires. Le premier point crée un phénomène de prévision homogènes par plage horaire avec l’impossibilité de revenir sur cette prévision une fois la plage horaire commencée selon le second point. Néanmoins les besoins peuvent varier en fonction des saisons du fait de la thermo sensibilité du marché, par exemple la marge du soir disparait en été.

Une troisième piste de classification est l’existence des jours ouvrés. Beaucoup d’entreprise n’étant pas complétement automatisé ou nécessitant une forte surveillance, le travail humain a une forte influence sur la consommation à l’échelle nationale. Une grande partie du travail d’Alpiq est de classer le profil de consommation d’entreprise selon la base et les pics de consommations.

Enfin le marché présente des situations de tensions différents en fonction des spécificités techniques de chaque offre. Le type de réserve et la direction du marché peuvent être utilisé

Il est possible de conserver des caractéristiques qualitatives pour chaque instant comme :

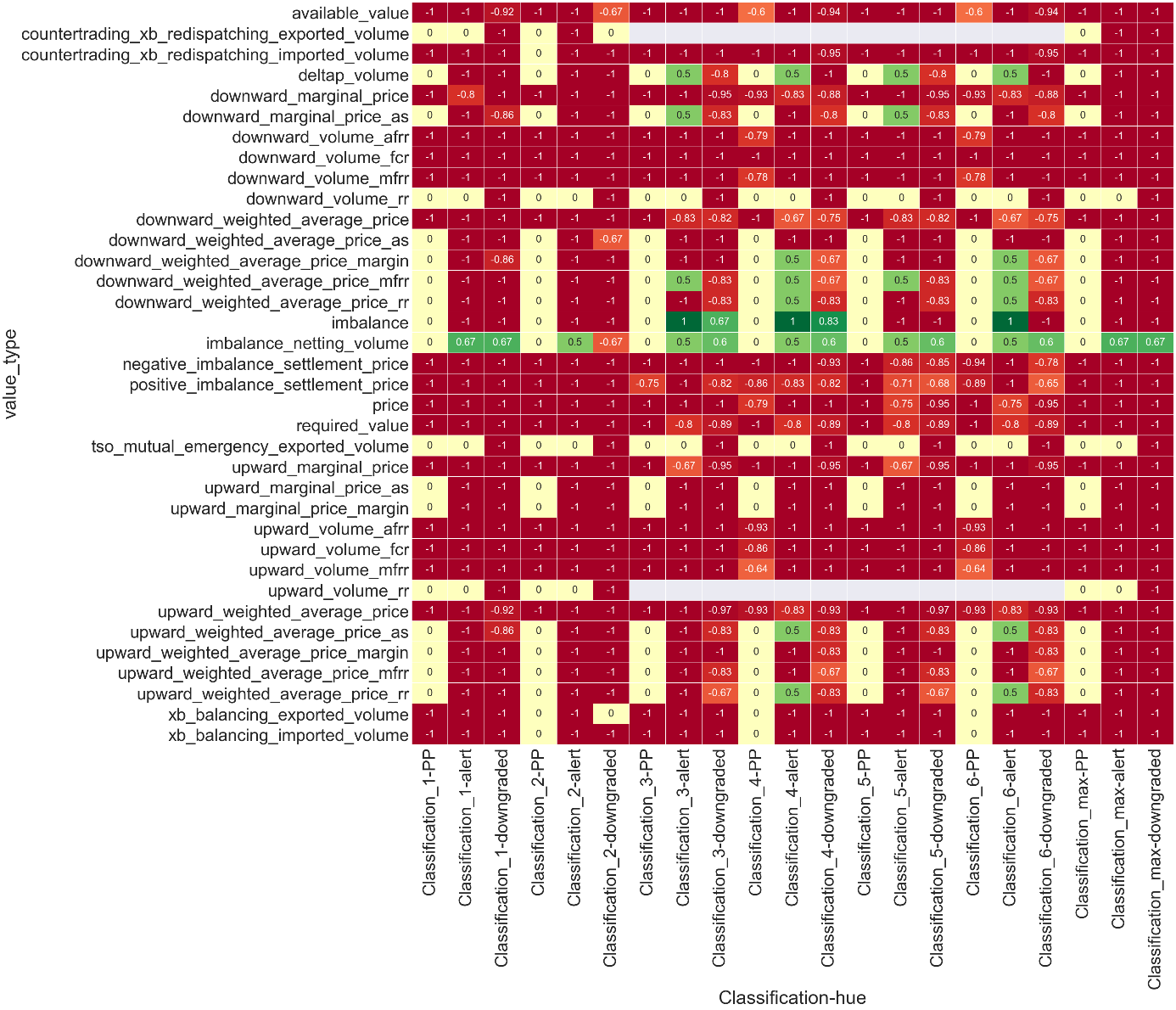
* Plage horaire
* Mois
* Jour ouvré
* Direction du marché ou de l’offre
* Type de réserves
* Type de tension (jours PP1/PP2, mode dégradé ou alerte)

Finalement, des résultats seront présentés pour les classifications suivante (la colonne de gauche présente la dénomination dans la suite de ce rapport des classifications correspondantes :

|  |  |
| --- | --- |
| Dénomination | Filtres de la classification |
| Classification\_1 | ["value\_type","month"] |
| Classification\_2 | ["value\_type","plage"] |
| Classification\_3 | ["value\_type","month","plage"] |
| Classification\_4 | ["value\_type","month","plage","bday"] |
| Classification\_5 | ["value\_type","month","plage","direction"] |
| Classification\_6 | ["value\_type","month","plage","bday","direction"] |
| Classification\_max | Classification obtenant la fréquence d’occurrences d’un événement la plus élevée |

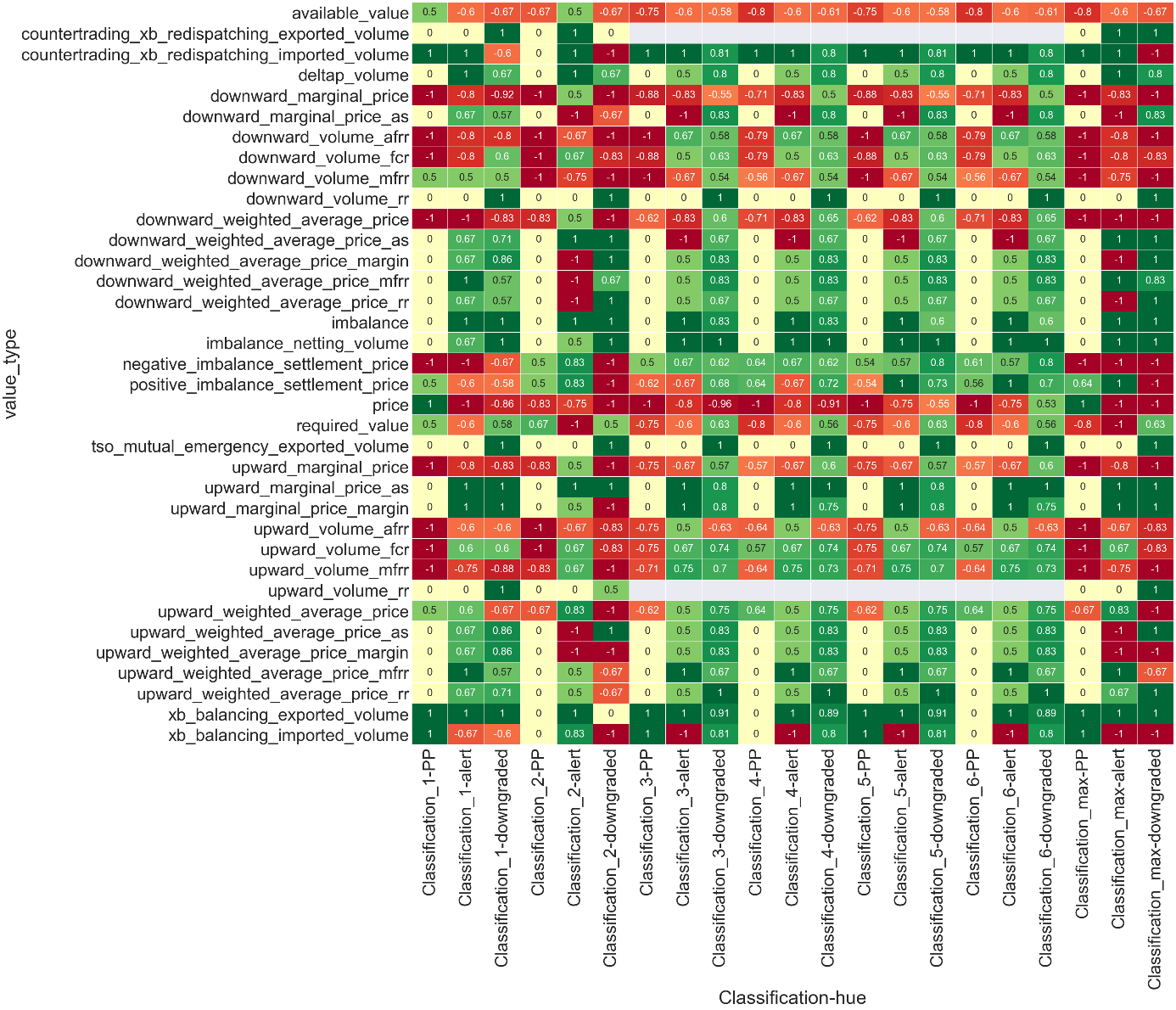
### Test de normalité

L’étude s’est faite à un risque 5%. C’est une valeur largement utilisée dans la littérature et qui correspond pour une loi normale à un écart positif ou négatif de deux fois l’écart type autour de la moyenne. Les résultats sont assez marqués et sans trop de contradiction. Pour toutes les ressources (sauf une), l’hypothèse nulle de normalité est rejetée quelle que soit la classification utilisée. Nous devrons donc utiliser des tests non paramétriques d’autant plus que les séries temporelles sont souvent dépareillées.

Seules une ressource, « imbalance\_netting\_volume », présentent possiblement une distribution normale (le test rejette la normalité mais ne la prouve pas) pour 5 classification sur 6. La 6ème classification présente un résultat opposé, l’hypothèse nulle est acceptée. Cela peut être rapprochée à une sorte d’« over-fitting », phénomène présent aussi pour les autres tests et que l’on peut comprendre par le nombre croissant de catégorie et donc une cardinalité de chaque catégorie décroissante. Cela peut aussi venir du fait que l’on essayé de deviner des classifications dans le but de trouver des distributions normales. Cette recherche est basée sur une connaissance des règles du marchés mais qui peut être qu’une vision communiquée par RTE de leur donnée pour éviter une utilisation non souhaitée.

La figure ci-dessus est un concentré des résultats de normalité dans une situation normale vis-à-vis des indicateur pris séparément. Pour la lire il est nécessaire de suivre l’annexe « Lecture des Résultats ». On observe le résultat précédemment expliqué, seul la ressource « imbalance\_netting\_volume » peut être considéré comme normal.

Le but de la classification était de trouver des sous-ensembles où la distribution était normale. Mais comme déjà mentionné cela peut mener à diminuer trop fortement la cardinalité de ces sous-ensembles. On observe sur la heatmap une polarisation plus forte sur les deux premières classification (1 et 2) que sur les suivantes. Lorsque l’on observe ces distributions avec les classifications 3, 4 ,5 et 6, il n’est pas trivial de trouver un schéma pour prédire de façon certaine la normalité des ensembles.

La figure ci-dessus est un concentré des résultats de normalité dans une situation de tension vis-à-vis des indicateur pris séparément. Contrairement à la situation normale, les résultats sont moins tranchés. La colonne max affiche un plus grand nombre d’acceptation de l’hypothèse nulle du test de normalité. Cela peut s’expliquer par des situations plus homogènes ou mieux partitionné par RTE laissant ainsi dans la situation de normalité un ensemble de situation différentes mais moins bien identifiées. Il serait intéressant d’analyser uniquement les instants non labelisé par RTE.

L’étude présente concerne la comparaison entre la situation normale et la situation de tension. Contenu des hypothèses reposant à la fois sur les tests paramétriques et non paramétriques, le choix non-paramétrique semble être le plus robuste pour continuer l’analyse.

### Analyse des impacts de la situation d’alerte sur les prix du Mécanisme d’ajustement

#### Lecture proposée des résultats :

Cette partie permet de comprendre la lecture des résultats qui sont peu visuels et qui empêche donc une représentation graphique.

##### Choisir la majorité

Lorsque les classifications de cardinalité plus faible indiquent un résultat contraire aux autres classifications. Mais que les secondes ont un résultat en valeur absolue proche de 0.5 on retiendra le résultat des premières.

##### Importance des filtres

Sachant que beaucoup de donnée contiennent intrinsèquement le filtre de marché à la hausse ou à la baisse (example : « downward\_marginal\_price »), le filtre « direction » n’aura pas tendance à être mis en avant en cas de balancement entre deux réponses.

##### Faire attention à l’over-fitting

On se concentre sur les feuilles « Clasification\_max, resultats\_kruskal et resultats\_mannwhitneyu ». L’idée générale déjà présentée en partie

**SI** une ressource présente une classification max avec une faible cardinalité avec un rejet de l’hypothèse nulle aux deux tests

**ET** des résultats positifs pour les tests de Kruskal-Wallis pour les classifications de forte cardinalité

**ET/OU** des résultats positifs pour les tests de Mann-Whitney pour les classifications de forte cardinalité

**ALORS** on pourra remettre en cause les résultats des tests de Kruskal-Wallis et/ou de Mann-Whitney et à mettre beaucoup de précaution sur les résultats d’appartenance à la même population.

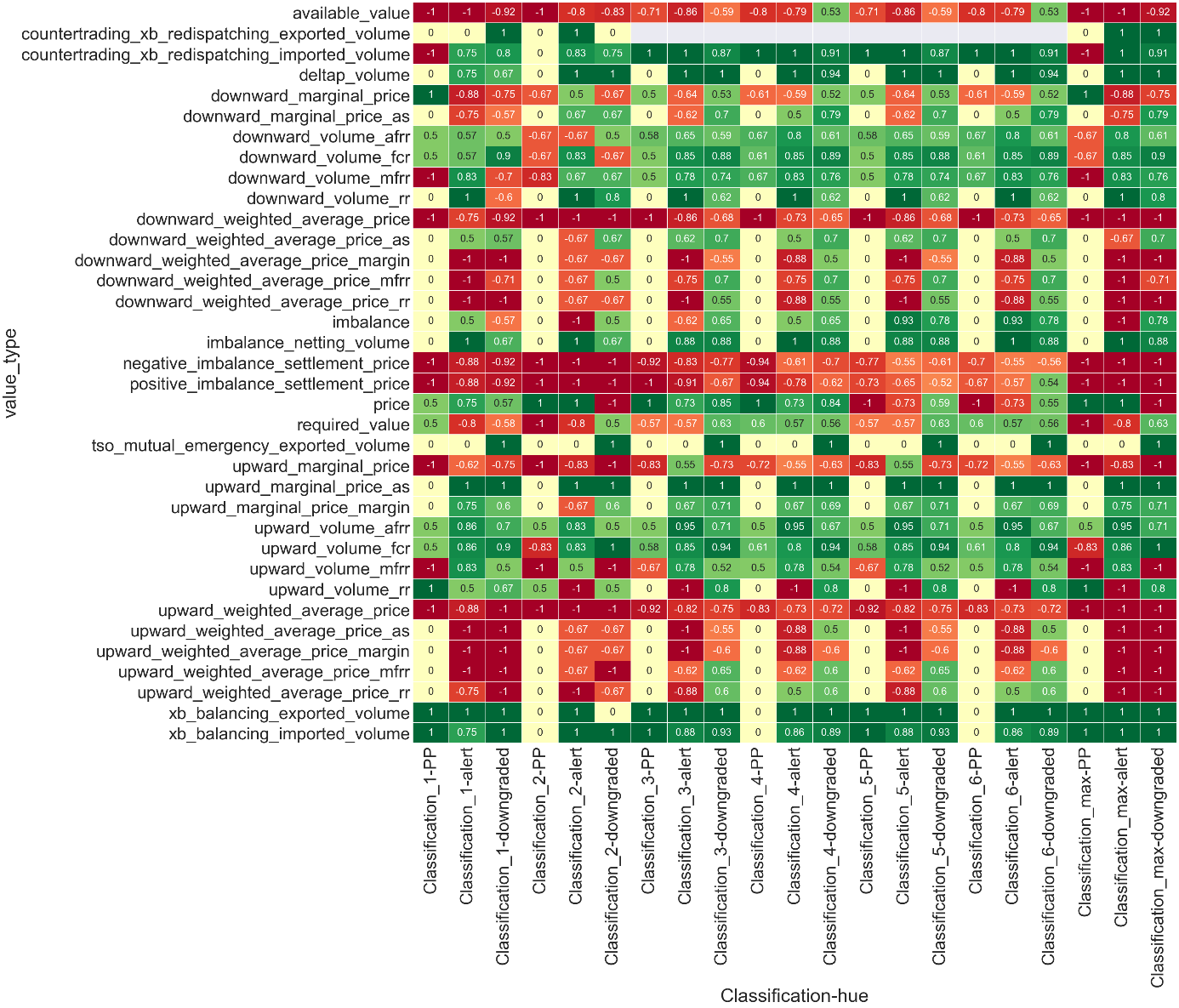
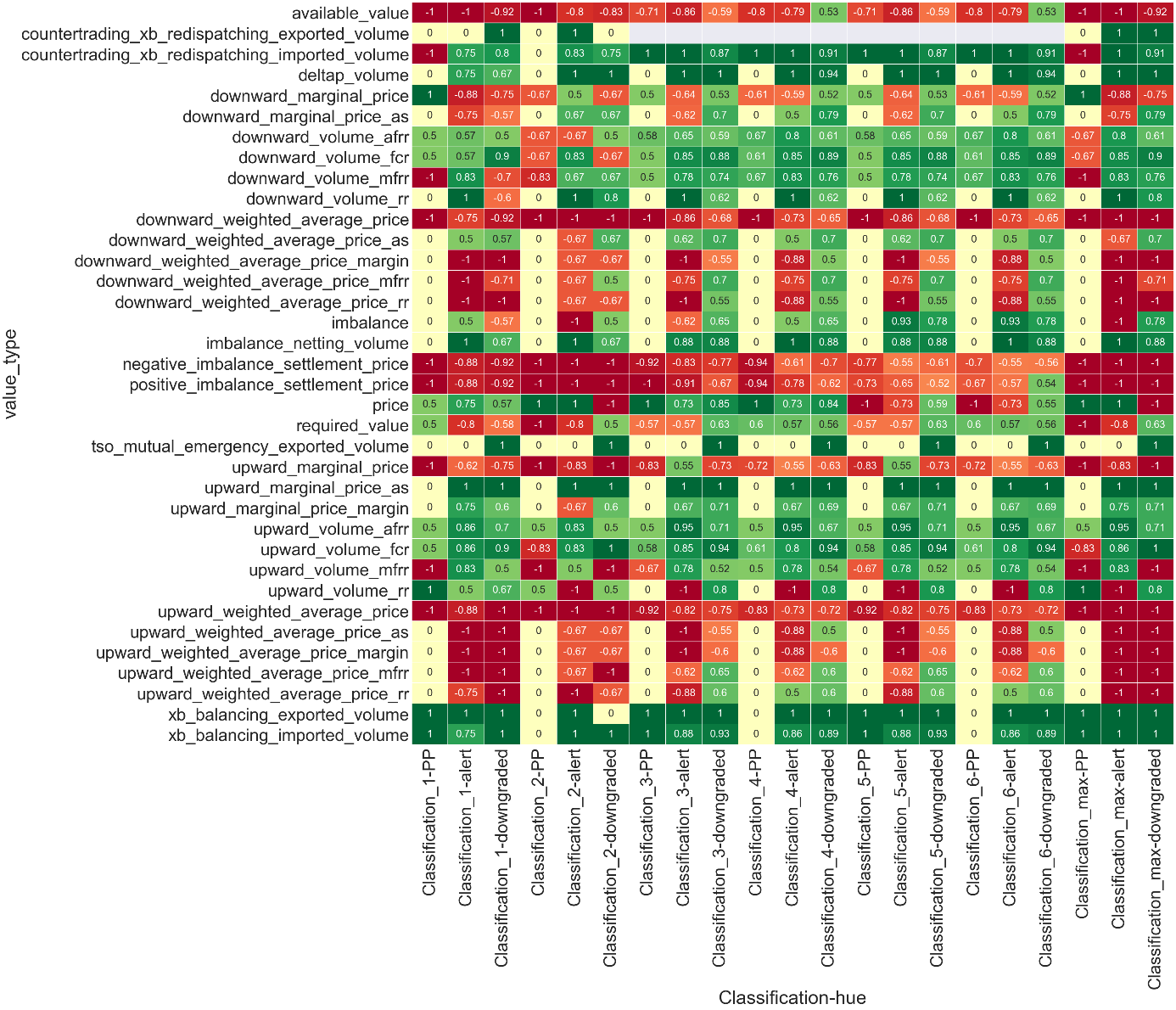
**FINALEMENT** Il sera préférable de conclure d’un rejet de l’hypothèse nulle

#### Schémas courants

##### Polarisation croissante

Les classifications sont classées par cardinalité croissante respectivement de gauche à droite. Néanmoins certaines différent par leurs filtres seulement. Il est possible d’observer des suites de gauches à droites non croissantes mais qui seront tout de même compté comme croissante. Dans l’idée de limité l’over-fitting on conservera le premier maximum rencontré de gauche vers la droite.

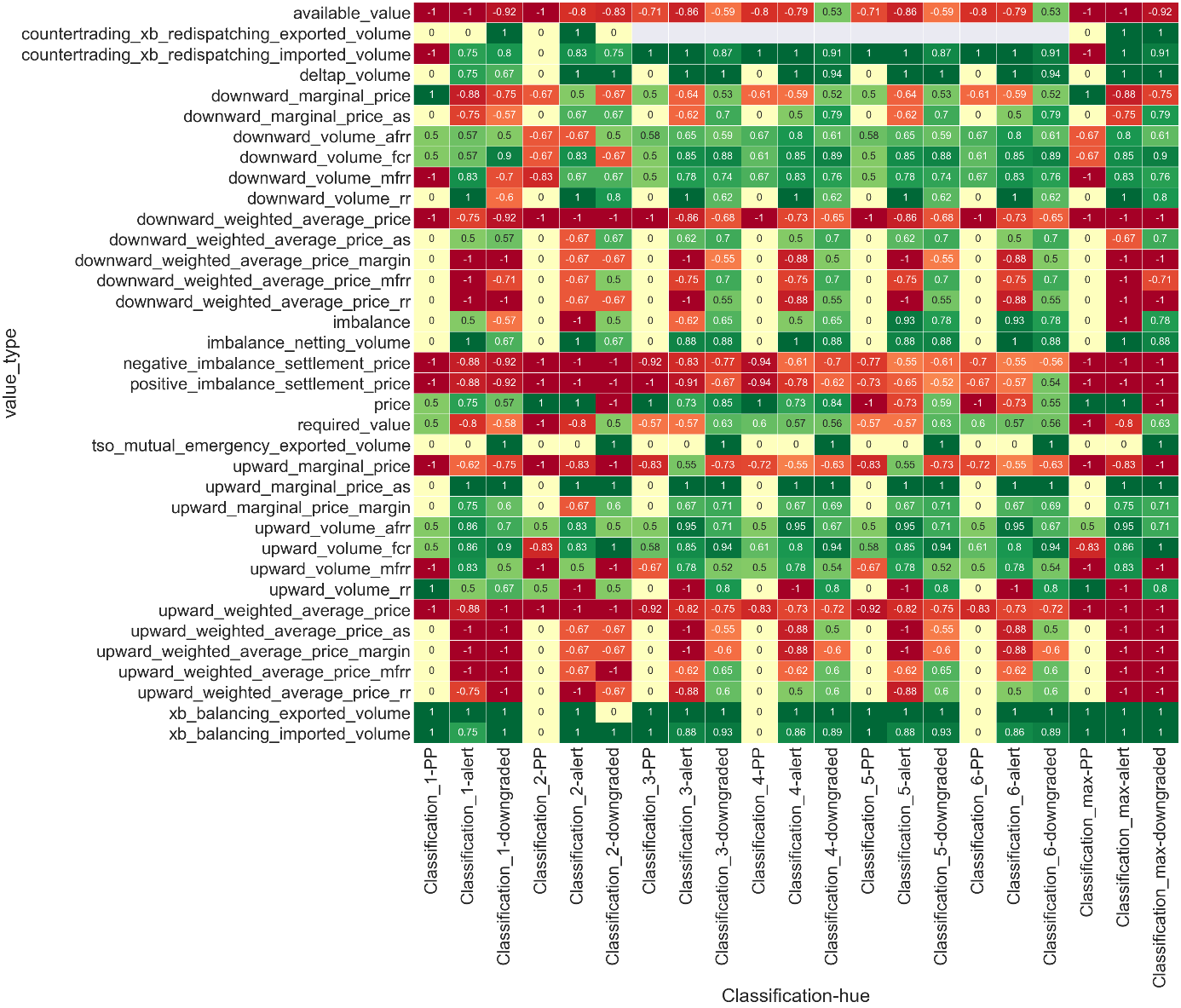
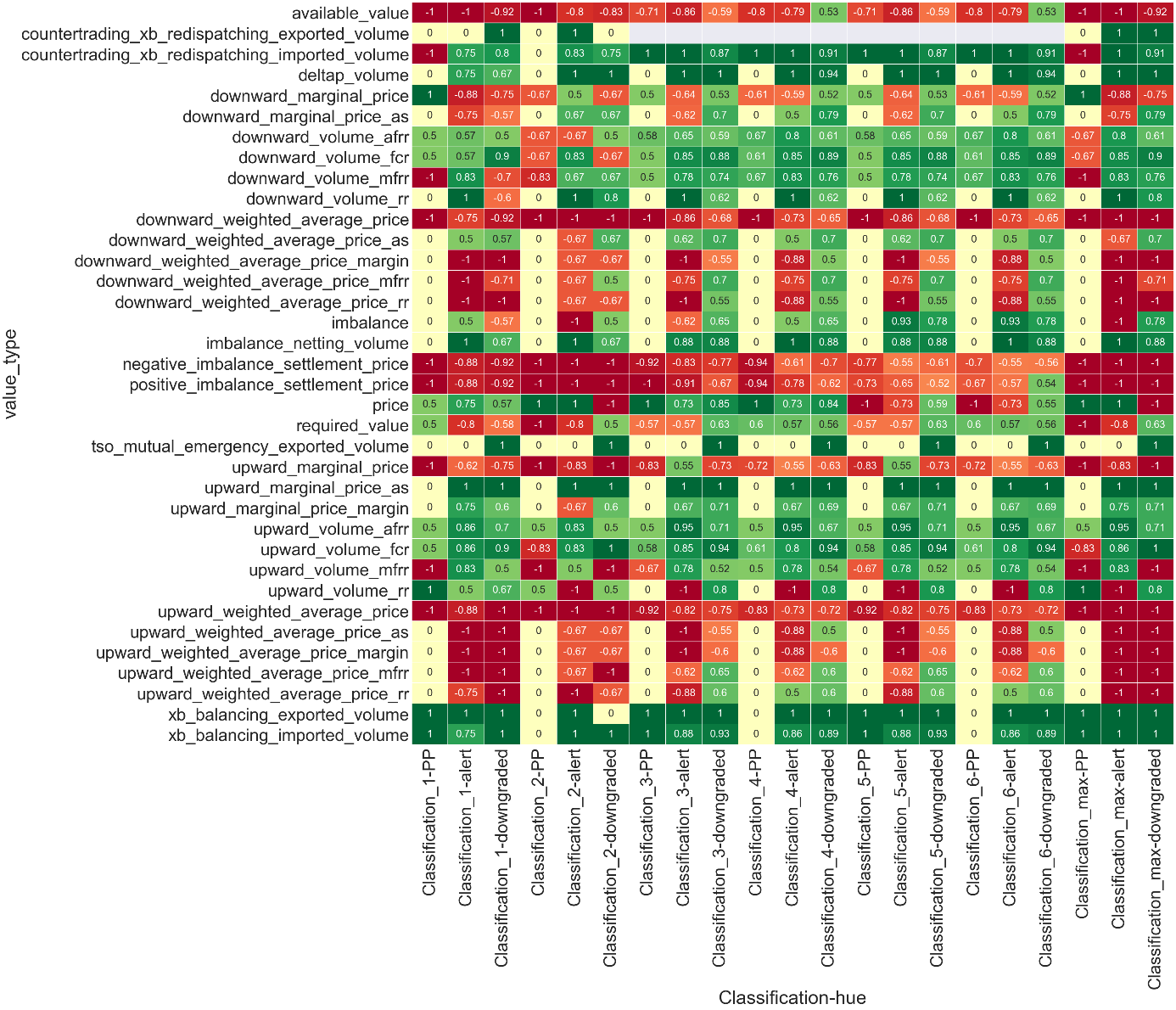
Exemple :



##### Polarisation constante

C’est la plus visuelle et la plus simple avec laquelle conclure. Quelle que soit la classification utilisée le test agrégé fourni toujours le même résultat. La réponse sera dans ce cas la pris dans la colonne max.

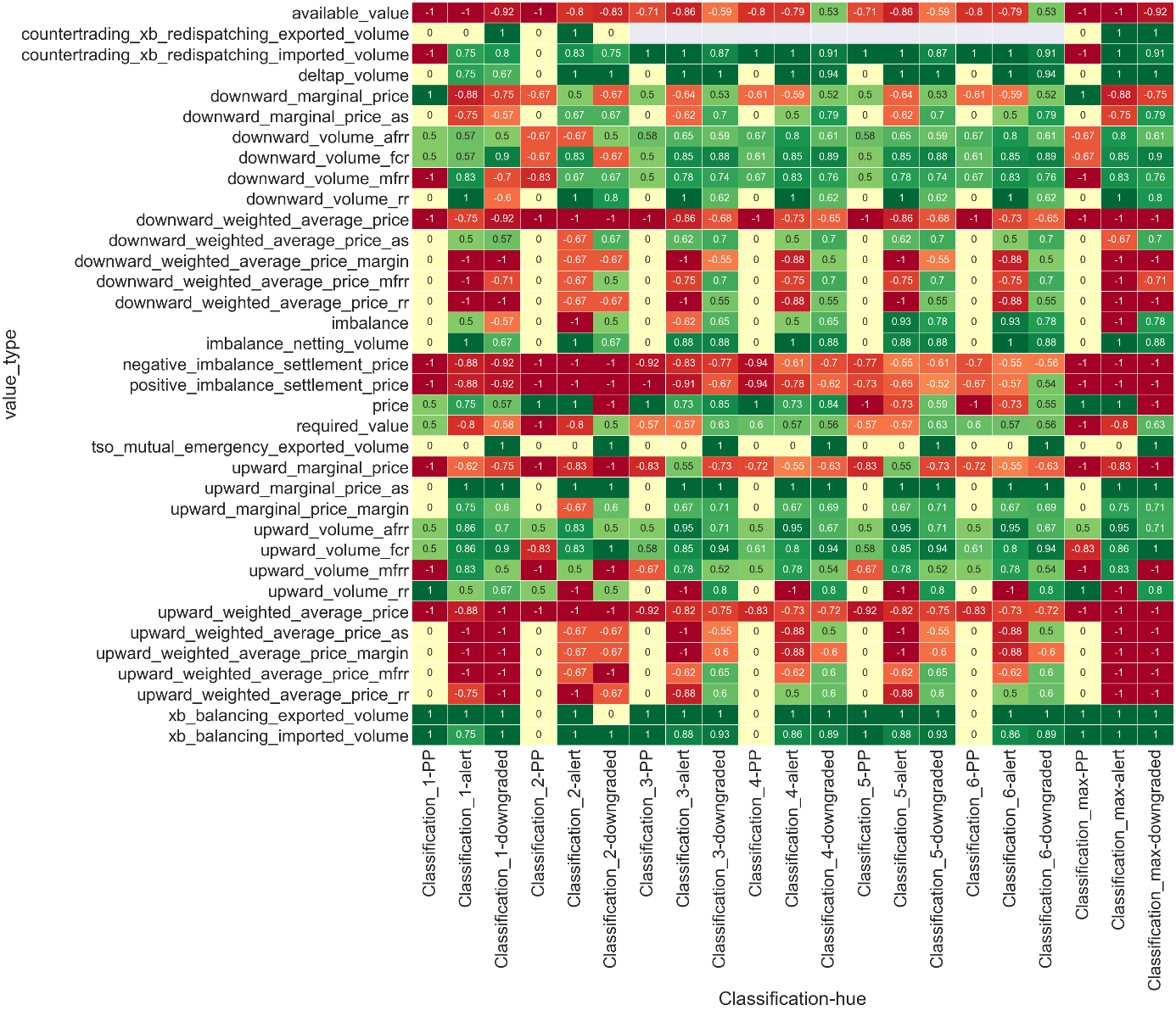
Exemple :



##### Polarisation décroissante

Les classifications sont classées par cardinalité croissante respectivement de gauche à droite. Sachant que le signe ne représente que l’issue du test, on peut sélectionner le ratio avec la valeur absolue de la case. Il est possible d’observer des suites de gauches à droites décroissantes non strictement. Dans l’idée de limité l’over-fitting on conservera le premier maximum rencontré de gauche vers la droite.

Exemple :

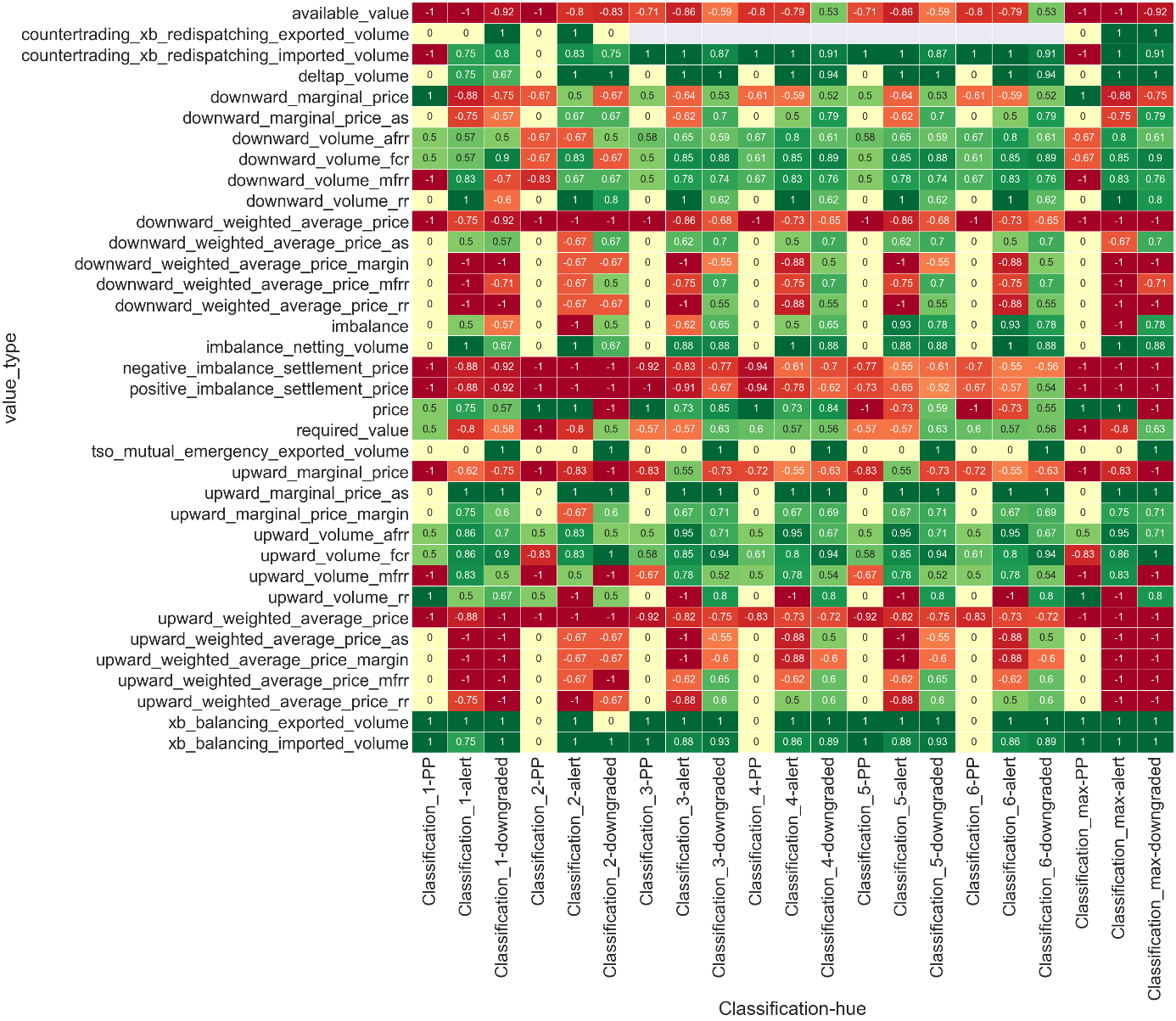
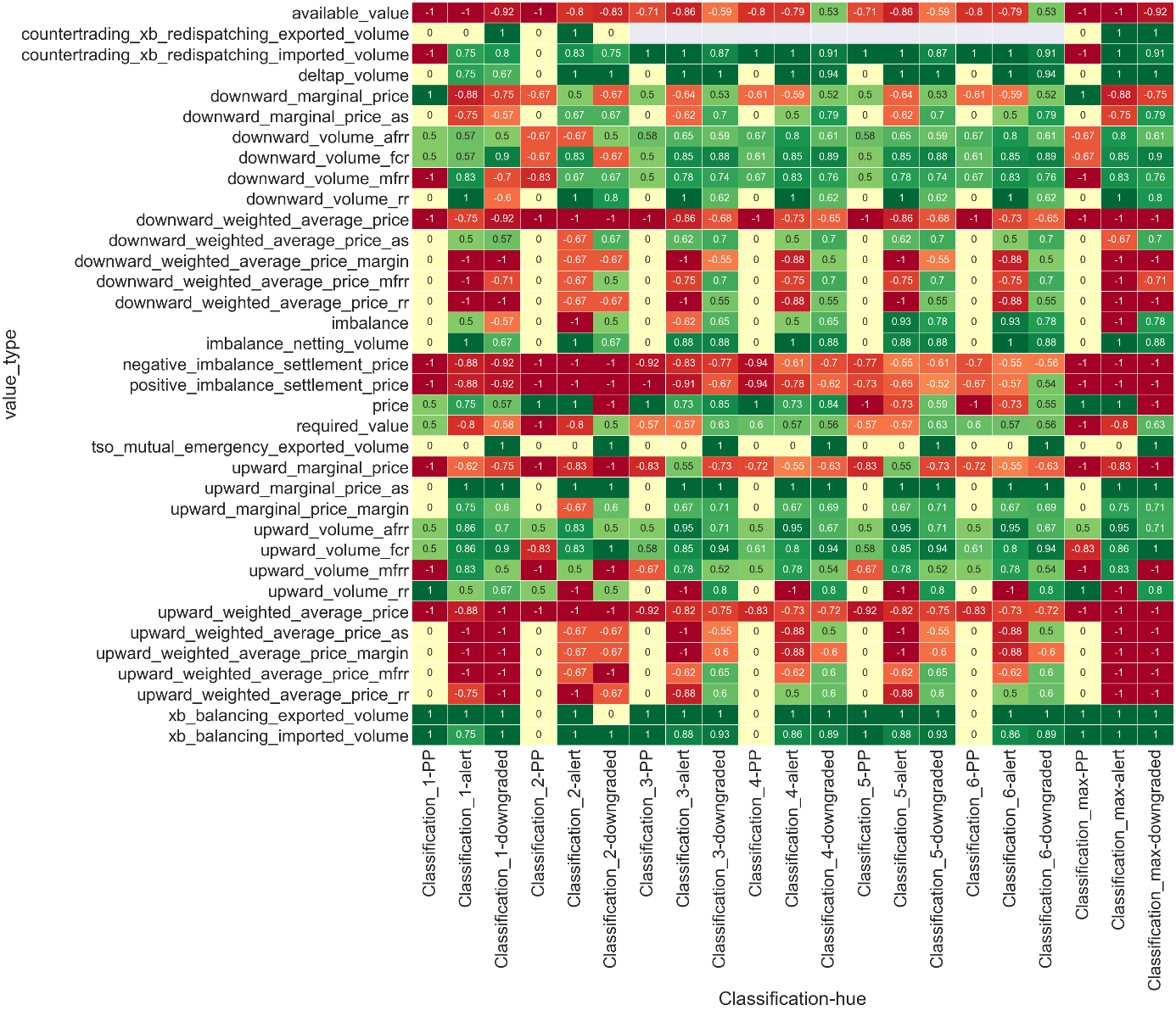


##### Polarisation en « V inversé »

Ce qui nous intéresse est d’être capable de sélectionner une réponse pour la ressource en générale. A partir de ce fait on peut donc comme cela nous arrange choisir de prendre la valeur telle qu’écrit dans la matrice soit son opposé. Quand la réponse sera le rejet, on propose de choisir l’opposé des valeurs de la matrice à la ligne de la ressource, et de prendre la valeur telle qu’elle quand la réponse est l’acceptation de l’hypothèse nulle. Il est donc toujours possible de se ramener à un « V inversé ».

Cette explication permet de retrouver le sommet du « V inversé » qui sera donc utilisé dans la colonne max.

Exemple :

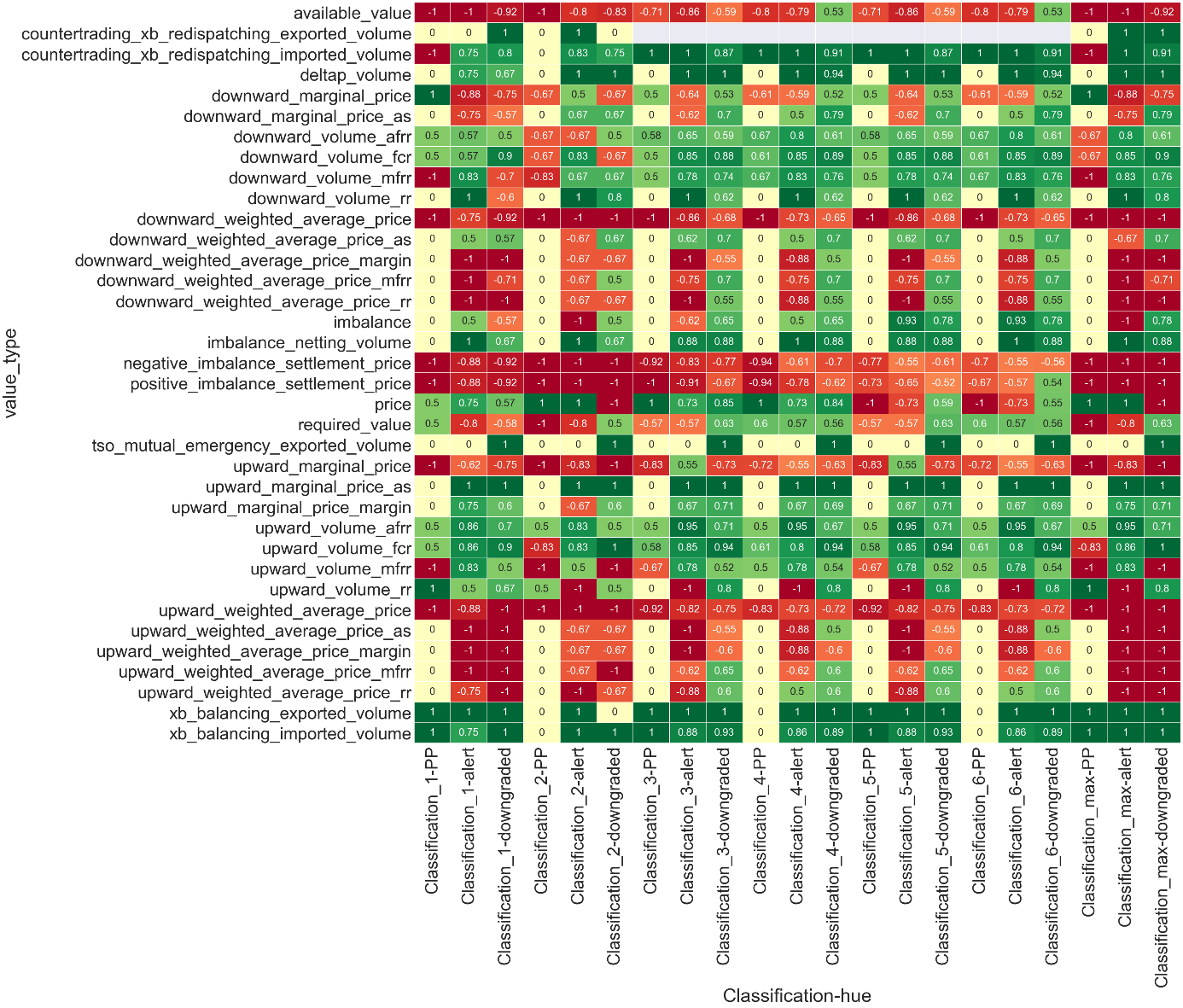
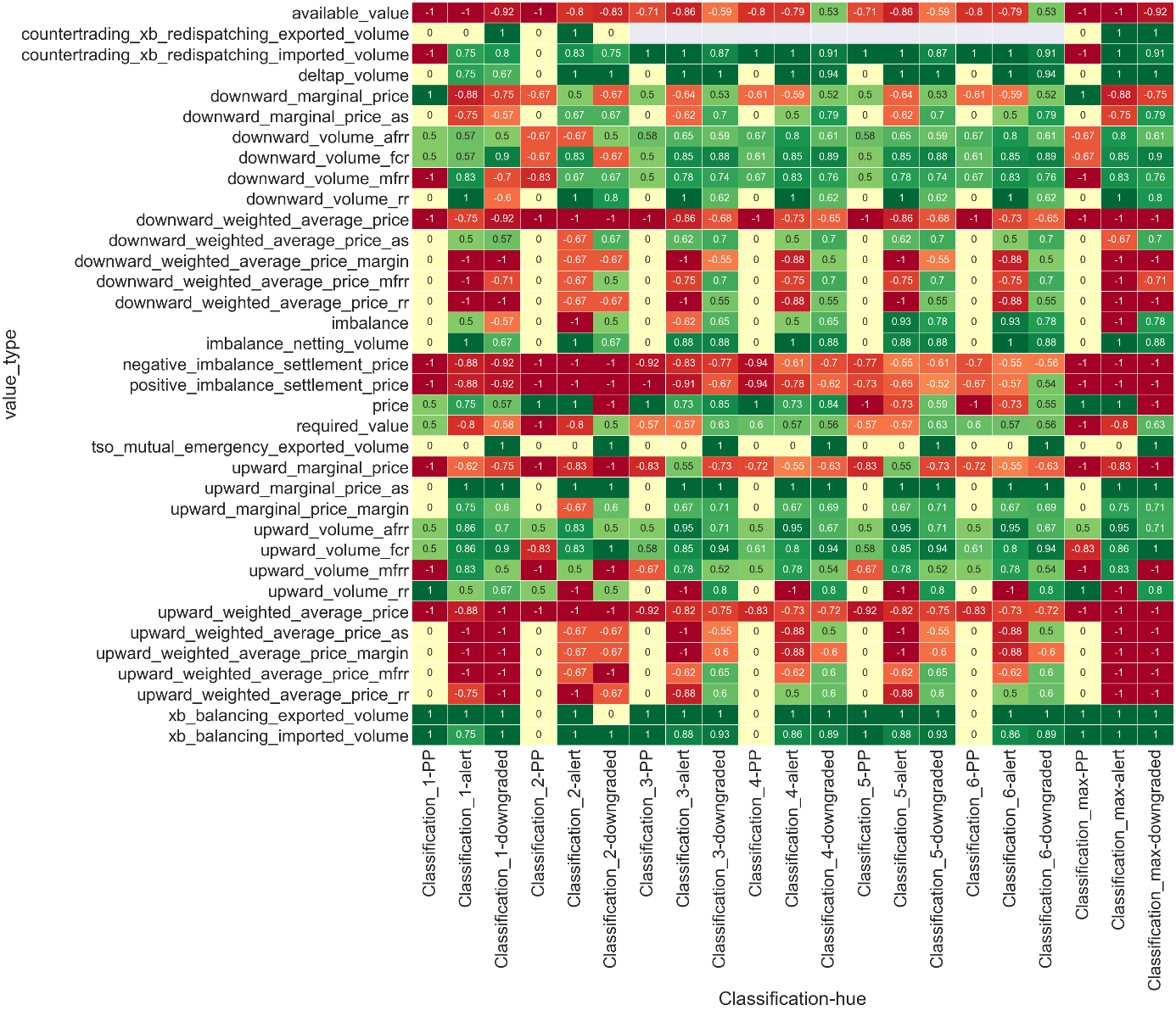


Ici la colonne max (c’est la conséquence, mais cela permet de comprendre le mécanisme) contient le rejet de l’hypothèse nulle, on devra lire pour chaque case noire l’opposé de la valeur écrite.

##### Polarisation avec « W »

C’est le schéma le moins clair et celui auquel on appliquera le plus les lectures précédemment définies. Les valeurs d’une ligne croisent plusieurs fois 0, il est difficile de définir quel est la tendance principale directement.

Example :



##### Conclusion sur les schémas

Pour les 4 premiers schéma l’utilisation de la colonne max et donc la prise du maximum de la ligne sera suffisante. Dans le dernier cas il sera nécessaire de faire un raisonnement en utilisant les lectures proposées.

#### Résultats du test de Kruskal-Wallis

|  |  |  |  |
| --- | --- | --- | --- |
| Ressource | Résultat (H0 : acceptée, H1 : rejetée) | | |
| PP | alert | downgraded |
| Available\_value | H1 | H1 | H1 |
| Countertrading\_xb\_redispatching\_exported\_volume |  | H0 | H0 |
| Countertrading\_xb\_redispatching\_imported\_volume | H0 | H0 | H0 |
| Deltap\_volume |  | H0 | H0 |
| Downward\_marginal\_price | H0 | H1 | H1 |
| Downward\_marginal\_price\_as |  |  |  |
| Downward\_volume\_afrr |  |  |  |
| Downward\_volume\_fcr |  | H0 | H0 |
| Downward\_volume\_mfrr | H1 | H0 | H0 |
| Downward\_volume\_rr |  | H0 | H0 |
| Downward\_weighted\_average\_price | H1 | H1 | H1 |
| Downward\_weighted\_average\_price\_as |  |  | H1 |
| Downward\_weighted\_average\_price\_margin |  | H1 | H1 |
| Downward\_weighted\_average\_price\_mfrr |  | H1 | H1 |
| Downward\_weighted\_average\_price\_rr |  | H1 | H1 |
| imbalance |  | H0 | H0 |
| Imbalance\_netting\_volume |  | H0 | H0 |
| Negative\_imbalance\_settlement\_price | H1 | H1 | H1 |
| Positive\_imbalance\_settlement\_price | H1 | H1 | H1 |
| Price | H0 |  |  |
| Required\_value | H1 | H1 |  |
| Tso\_mutual\_emergency\_exported\_volume |  |  | H1 |
| Upward\_marginal\_price | H1 | H1 | H1 |
| Upward\_marginal\_price\_as |  | H0 | H0 |
| Upward\_marginal\_price\_margin |  | H0 | H0 |
| Upward\_marginal\_price\_affr |  | H0 | H0 |
| Upward\_marginal\_price\_fcr |  | H0 | H0 |
| Upward\_marginal\_price\_mfrr | H1 | H0 | H1 |
| Upward\_marginal\_price\_rr | H0 | H1 | H0 |
| Upward\_weighted\_average\_price | H1 | H1 | H1 |
| Upward\_weighted\_average\_price\_as |  | H1 | H1 |
| Upward\_weighted\_average\_price\_margin |  | H1 | H1 |
| Upward\_weighted\_average\_price\_mfrr |  | H1 | H1 |
| Upward\_weighted\_average\_price\_rr |  | H1 | H1 |
| Xb\_balancing\_exported\_volume | H0 | H0 | H0 |
| Xb\_balancing\_imported\_volume | H0 | H0 | H0 |

Explications :

* Imbalance : Quand on trace la ressource selon les classifications 5 et 6 on obtient des distributions très bien formée en témoigne les tests de normalité précédents. Pour cette ressource nous mettrons de côté les problèmes d’over-fitting.

#### Résultats du test de Mann-Whithney

|  |  |  |  |
| --- | --- | --- | --- |
| Ressource | Résultat (H0 : acceptée, H1 : rejetée) | | |
| PP | alert | downgraded |
| Available\_value | H1 | H1 | H1 |
| Countertrading\_xb\_redispatching\_exported\_volume |  | H0 | H0 |
| Countertrading\_xb\_redispatching\_imported\_volume | H1 | H0 | H0 |
| Deltap\_volume |  | H0 | H0 |
| Downward\_marginal\_price | H1 | H1 | H1 |
| Downward\_marginal\_price\_as |  | H1 | H1 |
| Downward\_volume\_afrr | H1 | H1 | H1 |
| Downward\_volume\_fcr | H1 | H0 | H0 |
| Downward\_volume\_mfrr | H1 | H0 |  |
| Downward\_volume\_rr |  | H0 | H0 |
| Downward\_weighted\_average\_price | H1 | H1 | H1 |
| Downward\_weighted\_average\_price\_as |  | H1 | H1 |
| Downward\_weighted\_average\_price\_margin |  | H1 | H1 |
| Downward\_weighted\_average\_price\_mfrr |  | H1 | H1 |
| Downward\_weighted\_average\_price\_rr |  | H1 | H1 |
| imbalance |  | H1 | H1 |
| Imbalance\_netting\_volume |  | H0 | H0 |
| Negative\_imbalance\_settlement\_price | H1 | H1 | H1 |
| Positive\_imbalance\_settlement\_price | H1 | H1 | H1 |
| Price |  | H0 |  |
| Required\_value | H1 | H1 | H1 |
| Tso\_mutual\_emergency\_exported\_volume |  |  | H0 |
| Upward\_marginal\_price | H1 | H1 | H1 |
| Upward\_marginal\_price\_as |  | H0 | H0 |
| Upward\_marginal\_price\_margin |  | H0 | H0 |
| Upward\_marginal\_price\_affr |  | H0 | H0 |
| Upward\_marginal\_price\_fcr |  | H0 | H0 |
| Upward\_marginal\_price\_mfrr | H1 |  | H1 |
| Upward\_marginal\_price\_rr | H0 | H1 |  |
| Upward\_weighted\_average\_price | H1 | H1 | H1 |
| Upward\_weighted\_average\_price\_as |  | H1 | H1 |
| Upward\_weighted\_average\_price\_margin |  | H1 | H1 |
| Upward\_weighted\_average\_price\_mfrr |  | H1 | H1 |
| Upward\_weighted\_average\_price\_rr |  | H1 | H1 |
| Xb\_balancing\_exported\_volume | H0 | H0 | H0 |
| Xb\_balancing\_imported\_volume | H0 | H0 | H0 |

# Conclusion

# Références