



# Indice de perception de la corruption 2012

## Note technique sur la méthodologie

### Contexte

L'indice de perception de la corruption (IPC) a été mis en place en 1995 avec pour vocation d'être un indicateur composite utilisé pour mesurer la perception de la corruption dans le secteur public dans différents pays du monde. Durant les 17 dernières années, la méthodologie et les sources utilisées pour bâtir l'indice ont été ajustées et améliorées. Suivant un processus d'évaluation rigoureux, d'importants changements ont été apportés à la méthodologie en 2012. La méthode que nous employons pour regrouper différentes sources de données a été simplifiée et fait désormais appel à des données collectées sur un an seulement. Cette méthode va surtout nous permettre de pouvoir comparer les notes au fil du temps, ce qui n'était pas possible jusqu'alors. Etant donné les changements dans la méthodologie, il est important de noter que les notes obtenues par les pays en 2012 ne peuvent être comparées à celle de 2011 ou des éditions précédentes. La comparaison d'une année à l'autre ne sera possible qu'à partir de 2012.

### Méthodologie

La méthodologie suit quatre étapes de base : sélection des données sources, mise à l'échelle de ces données, regroupement des données remises à l'échelle, indication de la marge d'erreur.

#### **1. Sélection des sources de données**

L'IPC utilise un certain nombre de sources disponibles qui rendent compte de la perception de la corruption. Chaque source est évaluée à l'aune des critères énoncés ci-dessous. Chaque institution dont nous utilisons les données a été contactée afin de vérifier la méthodologie employée pour attribuer des notes et d'obtenir la permission de publier les notes remises à l'échelle provenant de chaque source, aux côtés de la note composite de l'indice.

- A) Un mode de collecte des données et une méthodologie fiables provenant d'une institution crédible :** nous devons pouvoir nous fier à la validité des données que nous utilisons. Chaque source doit donc provenir d'une institution professionnelle qui indique clairement la méthode qu'elle utilise pour collecter ses données. Cette méthode doit faire preuve d'une véritable rigueur scientifique. Par exemple, quand il est fait appel à « l'opinion d'un expert », nous nous assurons que cet expert est qualifié. Quand un sondage est réalisé auprès d'entreprises, nous vérifions que l'échantillon retenu est significatif.
- B) Des données qui traitent de la corruption dans le secteur public :** les questions posées ou l'analyse proposée doivent correspondre à la perception du niveau de corruption dans le secteur public – ce dernier point doit être explicite. Les questions peuvent traiter d'un type spécifique de corruption (la petite corruption, par exemple), ou de l'efficacité de la prévention de la corruption – dans les cas où cet angle d'approche est pertinent – dans la mesure où cela peut être un moyen de connaître le degré de corruption perçue dans le pays.
- C) Granularité quantitative :** l'échelle qu'utilise une source de données doit permettre une différenciation suffisante des données (un échelonnement sur quatre niveaux

minimum) sur le degré de corruption perçu dans le pays, de manière que ces données puissent être adaptées à l'échelle de l'IPC, qui va de 0 à 100.

**D) Comparabilité entre pays :** dans la mesure où l'IPC classe les pays les uns vis-à-vis des autres, les données sources doivent pouvoir être légitimement comparables avec celles d'autres pays. Elles ne doivent pas être spécifiques à un seul pays. La source de données doit donc mesurer la même chose dans chaque pays et lui attribuer une note selon la même échelle.

**E) Comparaison sur plusieurs années :** nous voulons pouvoir comparer la note d'un pays – et l'indice dans son ensemble – d'une année à l'autre. Les sources qui retranscrivent la perception de la corruption à un instant donné mais qui n'ont pas vocation à répéter l'expérience sont ainsi exclues de l'indice.

## 2. Standardisation des sources de données

Chaque source est ensuite standardisée pour devenir compatible avec les autres sources disponibles. Ce n'est qu'après cette étape que toutes les sources pourront être agrégées sur l'échelle de l'IPC. Cette standardisation consiste à convertir toutes les sources de données pour les rapporter à une échelle allant de 0 à 100 ; 0 représente le plus haut niveau de corruption perçue et 100 le plus bas niveau.

Toute source établie sur une échelle où la note la plus basse correspond au plus bas niveau de corruption doit d'abord être inversée. Pour cela, on multiplie chaque note de l'ensemble de données par -1.

Chaque note est ensuite standardisée (rapportée à une note z) en lui soustrayant la moyenne de l'ensemble des données puis en divisant le résultat par l'écart-type. L'ensemble de données ainsi obtenu est centré autour de zéro, avec un écart type de 0,5.

Pour que ces notes z soient comparables entre différents ensembles de données, il faut définir les paramètres de moyenne et d'écart-type comme paramètres généraux. Ainsi, quand un ensemble de données ne couvre qu'un nombre de pays limité, il est nécessaire d'imputer une note à chacun des pays qui ne sont pas inclus. Pour imputer une valeur aux pays manquants dans chaque ensemble de données, nous utilisons le logiciel statistique STATA, et plus particulièrement la commande « ICE ». Cette commande utilise des régressions multiples pour chaque ensemble de données afin de fournir une valeur estimée pour chaque pays faisant l'objet de données incomplètes dans chaque ensemble de données. La commande « ICE » exécute dix fois la fonction d'imputation multiple, générant dix valeurs estimées pour chaque note « manquante ». La moyenne et l'écart-type de l'ensemble de données sont calculés sur la moyenne des dix ensembles de données complets ; ils sont utilisés comme paramètre de standardisation des données brutes. Il est important de noter que les valeurs imputées sont uniquement utilisées pour générer ces paramètres et ne servent pas de données sources pour calculer les notes des pays figurant dans l'IPC.

Les notes z sont ensuite converties à l'échelle de l'IPC, qui va de 0 à 100. On utilise pour cela une formule de remise à l'échelle simple, qui établit la valeur moyenne de l'ensemble de données standardisées à environ 45 et l'écart-type à environ 20. Toute note allant au-delà de 0 ou de 100 est réajustée.

Cette méthodologie étant utilisée pour la première fois cette année, 2012 constituera l'année repère. Dans les années à suivre (après 2012), les paramètres de moyenne et d'écart-type utilisés pour les calculs de standardisation seront les mêmes valeurs que celles générées pour l'indice de base de 2012. En utilisant les mêmes paramètres d'année en année, nous pouvons comparer au fil du temps les notes standardisées et mises à l'échelle. Quand de nouvelles sources entreront en compte dans l'indice, les calculs de mise à l'échelle leur permettront d'être en adéquation avec les paramètres de base établis en 2012, afin de refléter au mieux les changements au fil du temps. Il faudra pour cela estimer si la moyenne et l'écart-type auront globalement changé depuis 2012, puis utiliser ces nouvelles valeurs –

qui pourraient avoir dévié de 50 et de 20 - pour remettre le nouvel ensemble de données à l'échelle.

### **3. Rassembler les données remises à l'échelle**

La note de chaque pays dans l'IPC est calculée comme une simple moyenne des toutes les notes réadaptées disponibles pour ce pays (il convient de noter que nous n'utilisons aucune des valeurs imputées comme une note comptant dans cette moyenne). Un pays ne se verra attribuer une note que si au moins trois sources de données sont disponibles pour calculer cette moyenne.

### **4. Indiquer la marge d'erreur**

Aux côtés de la note de l'IPC sont indiqués l'erreur type et l'intervalle de confiance, qui reflètent la variance de la valeur des données sources qui constituent la note de l'IPC.

L'erreur type est calculée en divisant l'écart-type des données sources remises à l'échelle par la racine carrée du nombre de sources. En utilisant l'erreur type, il est possible de calculer un intervalle de confiance de 90 % permettant de conclure à une distribution normale.

- FIN -