

# Analyse en composante principales d'un tableau de distributions macroéconomiques.

Sun Makosso-Kallyth

MDC Pain Centre, McMaster University 1280 Main Street West  
Hamilton, Ontario L8S 4K1 Canada.

makossok@macmaster.ca

**Résumé.** Cet article présente l'application de deux extensions de l'analyse en composantes principales à un tableau de distributions macroéconomiques. Les méthodes présentées s'inscrivent dans le cadre de l'analyse des données symboliques. Elles étendent l'analyse en composantes principales aux variables symboliques de type histogramme. Dans la première méthode on détermine les moyennes des histogrammes, on effectue ensuite une ACP classique du tableau des moyennes et on projette en éléments supplémentaires les hypercubes obtenus à partir de la transformation des histogrammes en intervalles. En outre, pour améliorer le codage des modalités des variables de type histogramme, on présente un nouveau codage basé sur les scores de Redit. Dans la seconde méthode en revanche, on détermine les quantiles, on définit par la suite une mesure de corrélation d'histogrammes à partir des quantiles qui se correspondent. On procède ensuite par la détermination des vecteurs propres de la matrice de corrélation et la projection des quantiles en éléments supplémentaires. Nous appliquons les deux méthodes à des données macroéconomiques et montrons leur intérêt en analyse des données exploratoire.

**Mots clés :** variable symbolique de type histogramme, analyse en composantes principales, quantiles, scores de Redit.

## 1 Introduction

L'objet de cet article est celui de comparer l'application de deux extensions de l'analyse en composantes principales (ACP) à un tableau de distributions macroéconomiques. L'objet de l'ACP d'un tableau de distributions dans un tel contexte est celui d'expliquer les relations entre des variables qui décrivent de grands agrégats économiques. On peut par exemple s'intéresser à la description du profil démographique et environnemental (niveaux d'émission de Gaz à effet de serre, nombre d'enfants par femme, etc.) de certaines zones géographiques telles que l'Afrique, l'Amérique du Nord et l'Europe. Dans une telle optique, si on souhaite comprendre les disparités socioéconomiques entre ces zones géographiques, on peut recourir à l'analyse de données symboliques. En effet, les récentes contributions en analyse des données symboliques ont permis d'énormes avancées en analyse des données exploratoires. Suite aux travaux de (Diday (1988) et Diday (1989)), l'analyse des données symboliques (ADS) a connu d'importants bouleversements au point de devenir une discipline à part entière en apprentissage