

GREDI
Groupe de Recherche en Économie
et Développement International



Cahier de Recherche / Working Paper
16-07

**L'impact de l'écotourisme sur le bien-être local dans les
zones protégées du Népal : une approche multiniveaux**

Marie-Eve YERGEAU

L'impact de l'écotourisme sur le bien-être local dans les zones protégées du Népal : une approche multiniveaux

Version préliminaire - Merci de ne pas citer

Marie-Eve Yergeau *

6 novembre 2016

Résumé

L'écotourisme dans les zones protégées est souvent cité comme étant un mécanisme contribuant à l'amélioration du bien-être local et à la réduction de la pauvreté lorsque les populations sont contraintes au niveau de l'utilisation des ressources naturelles. Cependant, l'effet direct sur le bien-être est encore peu étudié dans la littérature. Dans cet article, nous estimons l'effet de l'écotourisme sur le bien-être des populations locales dans les zones protégées du Népal. De plus, nous vérifions si les restrictions environnementales affectent le bien-être des ménages qui affirment être contraints au niveau de l'utilisation des ressources naturelles. Nous utilisons un modèle hiérarchique linéaire à deux niveaux afin de tenir compte de la structure de la base de données ainsi que de la présence de clusters. Nous estimons que l'implication d'un ménage dans le secteur écotouristique, en tant que travailleur autonome, affecte positivement et significativement son bien-être. De plus, nous observons que le secteur écotouristique génère une externalité positive sur le bien-être des ménages appartenant à une même communauté. Enfin, nous trouvons que le fait d'être contraint par la réglementation environnementale ne semble pas avoir un effet direct sur le bien-être local.

Mots clés : Bien-être, conservation, écotourisme, modèle multiniveaux, Népal

Codes JEL : I31 ; Q26 ; O13

*GRÉDI et LAMETA - Département d'économique - Université de Sherbrooke et Université de Montpellier - courriel : marie-eve.yergeau@usherbrooke.ca. L'auteure remercie le Centre de recherches pour le développement international ainsi que le Groupe de recherche en économie et développement international de l'Université de Sherbrooke pour leur soutien financier ayant permis la réalisation de la collecte des données utilisées dans cet article. Merci également à Dorothee Boccanfuso, professeure et chercheure du GRÉDI, département d'économique, Université de Sherbrooke pour son appui, son implication et ses commentaires, et à Stéphane Mussard, professeur et chercheur au CHROME, Université de Nîmes et chercheur associé au GRÉDI, pour ses commentaires et suggestions.

1 Introduction

Conservation environnementale et réduction de la pauvreté sont deux préoccupations majeures dans les pays en développement. Toutefois, la nature de la relation entre ces enjeux demeure l'objet de débats dans la littérature scientifique et politique (Brockington et Wilkie, 2015). D'une part, la pauvreté est perçue comme étant associée à une utilisation non-durable des ressources naturelles (BLI, 2016). Par exemple, on estime que l'agriculture de subsistance est la principale cause de déforestation dans les forêts tropicales (Lanly, 2003) et que 74 % des pauvres sont touchés par la dégradation des terres (UN, 2016). D'autre part, les politiques de conservation environnementale dans les pays en développement sont parfois critiquées sous prétexte qu'elles restreignent l'accès aux ressources, qui constituent souvent la source principale de revenu des populations les plus pauvres (Brockington et Wilkie, 2015). Malgré tout, la mise en place de zones protégées pour la conservation de la biodiversité et des écosystèmes est une pratique répandue et croissante. En effet, entre 1990 et 2014, la surface protégée mondiale est passée de 13,4 millions km² à 32 millions km², couvrant ainsi près de 15 % de la surface terrestre (UNEP, 2014). Dans les pays en développement, les zones protégées sont souvent mises en place dans les espaces reculés où les taux de pauvreté sont plus élevés (Dudley et al., 2008). Dans ce contexte, des mécanismes permettant de compenser les pertes causées par les restrictions imposées sur l'utilisation des ressources naturelles sont développés et mis en place. Certains consistent à offrir une compensation monétaire directe pour les initiatives environnementales alors que d'autres visent plutôt à introduire les efforts de conservation dans une stratégie de développement globale. L'écotourisme est l'un de ces mécanismes ayant retenu beaucoup d'attention, et étant de plus en plus intégré aux stratégies de réduction de la pauvreté.

Dans cet article, nous étudions la relation entre zones protégées, écotourisme et bien-être, dans le but d'en dégager les implications politiques. Nous utilisons des données népalaises originales et collectées dans une perspective de répondre à cet objectif. Notre étude se décline en deux principales questions de recherche. Premièrement, nous vérifions si l'écotourisme constitue un mécanisme permettant d'améliorer le bien-être des ménages résidant à l'intérieur d'une zone protégée. Nous estimons d'abord l'effet individuel pour un ménage d'avoir une occupation directement liée à l'écotourisme. Puis, nous mesurons l'externalité produite par cette implication individuelle, sur le bien-être des ménages appartenant à une même communauté. Dans la littérature, peu d'auteurs ont mesuré objectivement l'effet de l'écotourisme sur le bien-être local dans les zones protégées.

En effet, certains suggèrent que la hausse du bien-être générée par la mise en place de zones protégées s'explique notamment par le développement de l'écotourisme, sans toutefois vérifier cette hypothèse. (Sims, 2010 ; Ferraro, Hanauer et Sims, 2011 ; Ferraro et Hanauer, 2011 ; Canavire-Bacarreza et Hanauer, 2013). A notre connaissance, seuls Ferraro et Hanauer (2014) montrent que l'écotourisme est un mécanisme à travers lequel la mise en place de zones protégées contribue à la réduction de la pauvreté, en utilisant comme unité d'analyse le secteur de recensement au Costa Rica. Notre étude apporte une nouvelle évidence quant à l'impact de l'écotourisme sur le bien-être des ménages résidant à l'intérieur des zones protégées. De plus, il s'agit à notre connaissance de la première analyse effectuée à partir d'une unité aussi désagrégée que le ménage et tenant compte des externalités produites par l'implication dans le secteur écotouristique, sur le bien-être.

Deuxièmement, nous examinons l'effet des restrictions imposées sur l'utilisation des ressources naturelles par les politiques de conservation, sur le bien-être local. Dans la littérature, les auteurs mesurant l'effet de la protection du territoire sur le bien-être utilisent souvent des méthodes d'appariement¹. Ils comparent ainsi le niveau de vie des ménages résidant à l'intérieur d'une zone protégée (ou à l'intérieur d'une superficie protégée donnée), à des ménages semblables résidant à l'extérieur de la zone, estimant l'effet moyen de la protection. Ces méthodes imposent implicitement l'hypothèse que tous les ménages associés à une même mesure de protection sont également affectés par les politiques de conservation. Dans cette étude, nous relâchons cette hypothèse. En effet, nous considérons que parmi les ménages résidant à l'intérieur d'une zone protégée, certains sont contraints dans leur utilisation des ressources naturelles alors que d'autres ne le sont pas. L'effet des restrictions environnementales sur le bien-être des ménages contraints est donc susceptible de différer de celui sur le bien-être des ménages non-contraints. Ainsi, nous estimons l'effet pour un ménage d'être contraint au niveau de l'utilisation des ressources naturelles, sur son bien-être. Cette question est importante puisqu'un effet significatif indiquerait la nécessité de mettre en place des mécanismes de compensation, ou d'en améliorer le ciblage vers les ménages les plus contraints. Il s'agit à notre connaissance de la première estimation de cette relation basée sur des données ménages. Mentionnons enfin que nous ne visons pas à comparer le bien-être des ménages appartenant à une zone protégée à celui de ménages vivant à l'extérieur de la zone. Dans cette étude, nous examinons plutôt un ensemble de ménages appartenant à un territoire protégé et soumis à

1. A l'exception de Sims (2010) qui utilise des méthodes de régression. Voir par exemple Andam et al., 2010 ; Ferraro et Hanauer, 2011 ; Canavire-Bacarreza et Hanauer, 2013 ; Robalino et Villalobos-Fiatt, 2015.

des restrictions environnementales afin de mieux comprendre les écarts de bien-être existant entre eux.

Pour atteindre notre objectif, nous utilisons un modèle hiérarchique linéaire à deux niveaux. D'un point de vue méthodologique, l'utilisation de l'approche multiniveaux se justifie par la structure de la base de données utilisée. En effet, la modélisation multiniveaux est appropriée pour l'analyse de données clustérisées ayant une structure hiérarchique. De plus, cette approche permet de combiner les relations modélisées entre des variables mesurées sur différents niveaux². Ainsi, une variable du niveau inférieur peut être expliquée par d'autres variables du même niveau de même que des niveaux supérieurs. Nos résultats montrent que l'implication dans une occupation directement liée à l'écotourisme en tant que travailleur autonome génère un effet positif et significatif sur le bien-être alors que l'effet sur le bien-être des ménages impliqués en tant que salariés n'est pas significatif. De plus, l'augmentation de la part des travailleurs autonomes impliqués dans l'écotourisme à l'intérieur d'une communauté affecte positivement et significativement le bien-être des ménages appartenant à cette communauté, indépendamment de leur implication individuelle. Enfin, nous observons qu'être contraint au niveau de l'utilisation des ressources naturelles n'a pas d'effet significatif sur le bien-être.

L'article est divisé comme suit. Dans la prochaine section, nous discutons des études causales entre zones protégées, écotourisme et bien-être et présentons le cas du Népal. Ensuite, nous détaillons la méthodologie ainsi que les variables utilisées. Puis, le modèle estimé et les résultats obtenus sont présentés dans la section 5. Nous terminons par une discussion sur les résultats ainsi que leurs implications en termes de recommandations politiques.

2 Zones protégées, écotourisme et bien-être

2.1 Les études causales

Les études mesurant l'effet de la conservation environnementale sur le bien-être et la réduction de la pauvreté sont peu nombreuses, en particulier dans les pays en développement. D'une part,

2. Par exemple, une base de données dans laquelle des ménages appartiennent à un village, et le village appartient à un district, présente une structure hiérarchique à trois niveaux : le niveau 1 étant le ménage, le niveau 2 étant le village et le niveau 3 étant le district.

les analyses théoriques montrent généralement que les politiques environnementales contraignent l'utilisation optimale de la terre, générant une baisse du bien-être (ex. Robalino, 2007 ; Robinson et al., 2008 ; Robinson et Lokina, 2011). D'autre part, des études empiriques récentes ont montré que la mise en place de zones protégées a contribué au développement économique et à la réduction de la pauvreté au Costa Rica, en Thaïlande et en Bolivie (Andam et al., 2010 ; Sims, 2010 ; Ferraro et Hanauer, 2011 ; Ferraro, Hanauer et Sims, 2011 ; Canavire-Bacarreza et Hanauer, 2013 ; Robalino et Villalobos-Fiatt, 2015). Cependant, ces études n'expliquent pas *pourquoi* les zones protégées génèrent une réduction de la pauvreté. En effet, les auteurs suggèrent certains mécanismes pouvant expliquer leurs résultats, tels que la migration (Sims, 2010), le développement d'infrastructures (Ferraro et Hanauer, 2011 ; Canavire-Bacarreza et Hanauer, 2013) et l'écotourisme (Sims, 2010 ; Ferraro, Hanauer et Sims, 2011 ; Ferraro et Hanauer, 2011 ; Richardson et al. 2012 ; Canavire-Bacarreza et Hanauer, 2013). Cependant, la disponibilité de données appropriées et les difficultés méthodologiques font en sorte que, malgré leur importance d'un point de vue politique, les études quantitatives sur les mécanismes à travers lesquels les politiques environnementales affectent le bien-être sont rares (Coad et al., 2008 ; Ferraro et Hanauer, 2014). A notre connaissance, seuls Ferraro et Hanauer (2014) mesurent objectivement l'impact de certains de ces mécanismes sur le bien-être, à partir de données costa ricaines. Ils attribuent près de la moitié de l'effet au développement de l'écotourisme. Dans ce contexte, il nous apparaît pertinent d'apporter une nouvelle évidence quant à l'impact de l'écotourisme sur le bien-être local, à l'intérieur des zones protégées.

Malgré la rareté des études causales relatives aux mécanismes liant conservation et bien-être, les coûts et bénéfices pour les individus résidant à l'intérieur des zones protégées sont documentés et discutés dans la littérature (Coad et al., 2008 ; Brockington et Wilkie, 2015). Une revue de ces coûts et bénéfices est proposée par Coad et al. (2008). L'expropriation et le déplacement des populations, le transfert de droits de propriété à l'État, l'accès restreint aux ressources de la forêt et les conflits humain-faune sauvage sont parmi les coûts associés à la conservation. Parmi les bénéfices, le maintien des sols et la prévention de catastrophes naturelles, l'accès à l'eau, la disponibilité des ressources de la forêt, le développement d'infrastructures ainsi que la poursuite de certaines traditions culturelles et religieuses sont cités par les auteurs. D'autres mécanismes associés à la conservation environnementale visent plutôt à générer un revenu alternatif pour les populations locales, notamment les paiements pour services environnementaux, les programmes de conservation et de développement intégrés et l'écotourisme. Cependant, aucun de ces mécanismes n'est

reconnu dans la littérature comme étant une solution "gagnant-gagnant" pour l'atteinte simultanée des objectifs de conservation environnementale et de réduction de la pauvreté³. Par ailleurs, des études ont montré que dans les pays en développement, le tourisme contribue à la croissance économique (Eugenio, Morales et Scarpa, 2004 ; Sequeira et Nunes, 2008), à l'emploi et qu'il est l'une des principales sources d'exportation (Neto, 2003 ; Yunis, 2004). De plus, l'écotourisme dans les zones protégées est un mécanisme souvent cité comme ayant le potentiel de générer un revenu local et d'améliorer le bien-être des populations (ex. Adams et al. 2004 ; Richardson et al., 2012 ; Ferraro et Hanauer, 2014 ; Yergeau, 2015). Il s'agit, selon Metcalfe (2003), d'un mécanisme efficace pour intégrer une utilisation durable des ressources naturelles à des projets de développement (Metcalfe, 2003 cité dans Coad et al., 2008). L'Organisation mondiale du tourisme (OMT) définit l'écotourisme comme "toutes les formes de tourisme basées sur la nature, dans lesquelles la motivation principale des touristes est l'observation et l'appréciation de la nature, de même que des cultures et des traditions qui prévalent dans les zones naturelles"(traduction de WTO, 2010). En 2012, les Nations Unies ont adopté une résolution reconnaissant le rôle de l'écotourisme pour la réduction de la pauvreté et la conservation environnementale, et appelant les États membres à adopter des politiques pour la promotion du tourisme (WTO, 2013). Aussi, le développement de l'écotourisme est de plus en plus intégré dans les stratégies de réduction de la pauvreté (Yunis, 2004 ; Goodwin, 2006 ; Chok, Macbeth et Warren, 2007). Puisque à ce jour, peu d'études empiriques mesurent objectivement l'impact de l'écotourisme sur le bien-être des populations locales (Meng, Li et Uysal, 2010), notre analyse vient en appui à ces recommandations.

2.2 Le cas du Népal

Malgré certains progrès réalisés au niveau de la croissance et du développement humain, le Népal demeure parmi les moins développés dans le monde. En 2014, 25,2 % de la population vivaient sous le seuil national de pauvreté (ADB, 2016). De plus, le Népal se classe au 145^{ième} rang sur 188 pays au niveau de l'indice de développement humain (UNDP, 2015b). L'agriculture est l'activité économique prépondérante dans le pays. En effet, le secteur, qui contribue à 35 % du PIB, prédomine au niveau de l'emploi, occupant 75 % de la main-d'oeuvre. Cependant, la majorité des

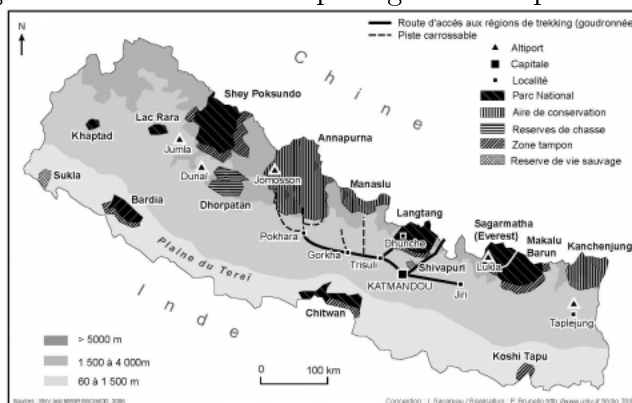
3. Pour plus de détails sur les mécanismes, voir Coad et al. (2008). Pour des études de cas analysant les impacts de ces différents mécanismes sur le bien-être ou la réduction de la pauvreté, voir par exemple Adhikari (2005) ; Grieg-Gran, Porras et Wunder (2005) ; Baral, Stern et Heinen (2006) ; Nagendra et Gokhale (2008).

fermiers sont pauvres (Basnet et al., 2014) : en région rurale, près de 80 % des ménages pratiquent l'agriculture de subsistance (IFAD, 2014). De plus, le rendement agricole est très inégal entre les différentes régions compte tenu des conditions géographiques irrégulières caractérisant le pays. En effet, le Népal se divise en trois régions écologiques : au sud, la région du Terai caractérisée par un sol plat et fertile, au centre, la région des Collines, et au nord, celle des Montagnes⁴. Ainsi, l'altitude varie entre 70 mètres au-dessus du niveau de la mer dans la région du Terai, et 8 848 mètres dans la région des Montagnes (Mont Everest). La région du Terai est propice à l'agriculture alors que dans les Montagnes, les conditions sont difficiles et les terres peu fertiles. Ces disparités géographiques se reflètent également au niveau des conditions de vie. En effet, en 2011, le taux de pauvreté dans les régions des Collines et du Terai était de moins de 25 % (respectivement 24,32 % et 23,44 %) alors qu'il était de 42,27 % dans la région des Montagnes (CBS, 2011). L'éloignement, la difficulté d'accès, les mauvais réseaux de communication et les infrastructures peu développées sont cités comme les facteurs contribuant à l'importance de la pauvreté en région montagneuse (IFAD, 2013).

Le Népal est également caractérisé par un environnement naturel riche mais aussi très fragile. Les défis environnementaux auxquels le pays doit faire face incluent la déforestation, la dégradation des terres, la perte de biodiversité, la fonte des glaciers et la pollution de l'eau et de l'air (RRN et CECI, 2007). En réponse à ces défis et aux différents objectifs de conservation et de développement, un système de zones protégées a progressivement été mis en place. En effet, la désignation des zones protégées au Népal a officiellement débuté en 1973 avec le passage du *National Parks and Wildlife Conservation Act*, afin de lutter contre la déforestation et le braconnage (Shrestha et al., 2010). La réglementation originale très stricte a ensuite été amendée afin que la désignation des zones intègre aux objectifs de conservation, le développement d'opportunités économiques durables (Keiter, 1995; Heinen et Shrestha, 2006). Aujourd'hui, les 20 zones protégées établies couvrent 23,23 % du territoire (Acharya, 2014). De ce nombre, on compte dix parcs nationaux, trois réserves fauniques, six aires de conservation, une réserve de chasse et douze zones tampons. La surface terrestre protégée est importante, de plus, il semble que plusieurs zones protégées aient atteint leurs objectifs de conservation (Shrestha et al., 2010). Les parcs nationaux ainsi que les réserves fauniques ont pour principal objectif la conservation de la biodiversité et des écosystèmes, et sont soumis à une réglementation plus stricte au niveau de l'extraction et de l'utilisation des ressources

4. La capitale Kathmandu se situe vers le centre du pays, dans la région des Collines

Figure 1. Carte des zones protégées au Népal en 2009



Source : SNV et MENRIS/ICIMOD dans Sacreau (2009)

naturelles. Les aires de conservation, réserve de chasse et zones tampons visent à promouvoir une utilisation durable des ressources, en intégrant le développement économique et social aux objectifs de conservation (IUCN, 2016). En fait, plus de 75 % des zones protégées au Népal visent à impliquer les populations locales dans la gestion des ressources et dans le partage des bénéfices de la conservation (Budhathoki, 2005).

Les richesses naturelles et l'héritage culturel du pays ont attiré de nombreux visiteurs depuis l'ouverture des frontières dans les années 50 (Nepal, 2000). Aujourd'hui, le Gouvernement du Népal considère le tourisme comme un contributeur majeur à l'économie (DNPWC, 2014). En 2014, 790 118 touristes sont entrés au pays (MCTA, 2015). De ce nombre, près de 55 % ont accédé à une zone protégée. La contribution totale⁵ du secteur touristique était alors évaluée à 8,9 % du PIB. De plus, l'industrie du tourisme a généré durant cette année 1 059 000 emplois, ce qui représentait 7,5 % de l'emploi total (WTTC, 2015). Dans les zones protégées, l'écotourisme est perçu comme ayant le potentiel d'améliorer les opportunités d'emploi non-agricole et de fournir des incitatifs à conserver les ressources (Mehta et Kellert, 1998). De plus, des études ont montré que les résidents ont une attitude positive envers l'écotourisme, en particulier ceux directement impliqués dans le secteur (Mehta et Kellert, 1998 ; Spiteri et Nepal, 2008). Enfin, Nepal (2000) soutient que l'écotourisme a permis à des régions reculées telles que celles de l'Everest et de l'Annapurna de sortir de la pauvreté. Cependant, des études mesurant quantitativement les effets directs de l'écotourisme sur le bien-être des populations locales dans les zones protégées manquent à la littérature.

Le 25 avril 2015, un séisme de magnitude 7.8 a frappé le Népal. Les dommages matériels incluant la destruction de maisons privées, de fermes, d'infrastructures publiques, d'entreprises et de mo-

5. Telle que définie dans WTTC, 2015 ; p.2

numents historiques, ont été évalués à 7,1 billions \$US. De plus, le gouvernement estime que le séisme a fait basculer près de 1 million de Népalais sous le seuil national de pauvreté (World Bank, 2016). L'impact sur l'industrie du tourisme fut immédiat et fort. Peu de statistiques officielles sont disponibles, mais le *Department of Immigration* estime à 554 747 le nombre de visiteurs en 2015, ce qui représente une baisse de 29,7 % par rapport à 2014 (Shrestha, 2016)⁶. Par ailleurs, moins d'un an plus tard, les arrivées touristiques étaient à nouveau en croissance, indiquant que le secteur a le potentiel de se rétablir suite à ce choc majeur. De nombreux articles de presse appellent les touristes à revenir, affirmant que le pays est sécuritaire et prêt à les recevoir. Dans cette perspective, l'analyse proposée dans cet article est d'autant plus pertinente puisqu'elle s'inscrit dans un contexte où une stratégie de reconstruction basée sur des secteurs favorisant un développement local durable devrait être réfléchi et mise en place.

3 Méthodologie

Dans cette section, nous détaillons la méthodologie utilisée. Nous présentons d'abord les données, puis la méthode économétrique, le choix du modèle s'appuyant sur la structure de la base de données. Nous discutons ensuite de la méthode d'estimation ainsi que des difficultés liées à l'utilisation de poids d'échantillonnage. Enfin, nous présentons la simulation effectuée afin de vérifier l'effet de différents facteurs d'échelle modifiant les poids d'échantillonnage sur les paramètres estimés, et les résultats obtenus.

3.1 Les données utilisées

Les données utilisées ont été collectées en 2013 dans le but spécifique de mesurer l'impact du développement écotouristique sur le bien-être des ménages résidant à l'intérieur des zones protégées du Népal, ce qui en fait une source de données parfaitement adaptée à la question de recherche. L'échantillon compte 1 563 ménages, sélectionnés selon un plan d'échantillonnage à plu-

6. Cette baisse de l'activité touristique peut également avoir été accentuée par un bouleversement politique avec l'Inde qui a eu lieu en septembre 2015, générant une pénurie d'électricité et d'essence. On estime à près de 45 % la diminution de visiteurs indiens en 2015, qui constituent par ailleurs le marché principal (Cette statistique n'est pas officielle et provient de <http://kathmandupost.ekantipur.com/news/2016-01-22/arrivals-hit-6-year-low-as-quake-agitation-take-toll.html>.)

sieurs degrés⁷. Trois zones protégées ont été considérées pour l'enquête : l'aire de conservation de l'Annapurna, le parc national de Langtang et le parc national de Chitwan. Cependant, la loi interdisant de s'établir à l'intérieur du parc national de Chitwan, les ménages résidant dans la zone tampon entourant le parc ont été interrogés. Ces zones ont été sélectionnées à partir de deux critères, soit l'année de désignation de la zone protégée et l'importance de l'activité touristique à l'intérieur de la zone, telle que mesurée par le nombre d'arrivées de touristes internationaux en 2012. Ces critères ont permis de sélectionner des zones caractérisées par une durée de protection et une activité touristique suffisamment importantes pour que les résidents interrogés puissent évaluer les impacts.

TABLE 1 – Description des zones protégées enquêtées

	Année de désignation de la zone ¹	Nombre de touristes en 2012 ²	Population totale (2011) ³	Nombre de VDC sélectionnés	Nombre de ménages enquêtés
Aire de conservation de l'Annapurna	1992	102 570	59 570	4	536
Parc National de Langtang	1976	14 315	17 619	3	491
Parc National de Chitwan	1973	130 364	93 334	3	536*

Sources : ¹DNPWC (2014) ; ²MTCA (2013) ; ³CBS(2012). *Ménages résidents de la zone tampon.

Les divisions administratives du Népal ont été utilisées afin d'élaborer un plan d'échantillonnage à plusieurs phases, qui a été répété à l'intérieur de chacune des trois zones protégées visitées. Premièrement, des *Village Development Committee* (VDC)⁸ ont été sélectionnés avec probabilité proportionnelle à la taille de la population (PPT) dans le VDC, telle que recensée par le Bureau Central de la Statistique du Népal en 2011. Puis, les VDC étant divisés en neuf *wards*⁹, le nombre de ménages par *ward* devant être interrogés a été déterminé de telle sorte que la probabilité d'être sélectionné d'un ménage soit proportionnelle à la taille de la population dans le *ward*. Enfin, les ménages à l'intérieur de chaque *ward* ont été sélectionnés par échantillonnage systématique¹⁰. Le tableau 7 en annexe A résume la structure de la base de données.

Le plan d'échantillonnage utilisé fait en sorte que les observations sont regroupées en cluster, sur différents niveaux hiérarchiques. En effet, les ménages (niveau 1) appartiennent à un *ward* (niveau 2), qui eux appartiennent à un VDC (niveau 3) (Figure 2 en annexe A). Aussi, les ménages appar-

7. Une fois pondéré, l'échantillon compte 170 157 ménages, ce qui est conforme à la population recensée en 2011 dans les trois zones protégées visitées, de 170 523 ménages (CBS, 2012).

8. Le VDC est une division administrative analogue à la municipalité.

9. Le *ward* est la plus petite division administrative au Népal.

10. Pour plus de détails sur le plan d'échantillonnage, voir A COMPLÉTER

tenant à un même *ward* et à un même VDC sont susceptibles d'être corrélées entre elles, violant l'hypothèse d'indépendance entre les observations. Les méthodes d'estimation standards, basées sur cette hypothèse, ne sont donc pas appropriées puisqu'elles peuvent mener à une sous-estimation importante des écart-types, et ainsi à conclure sur une significativité erronée des variables (Deaton, 1997 ; Cameron et Miller, 2010 ; Hox, 2010). Dans la littérature empirique, différentes méthodes sont proposées pour procéder à des estimations en présence de données clusterisées. Parmi celles-ci, une approche répandue consiste à utiliser une adaptation de la matrice de covariance de White (1980), telle que définie dans Cameron et Miller (2010, eqs. [6] et [7]), robuste à l'effet des clusters. Une autre méthode suggérée consiste à intégrer dans le modèle un effet fixe spécifique à chacun des clusters, afin de capter l'hétérogénéité entre les groupes. Cette méthode permet également de contrôler pour un biais de variables omises, dans le cas où les régresseurs du modèle sont corrélés à l'hétérogénéité entre les clusters (Cameron et Miller, 2010). Par ailleurs, une estimation convergente des paramètres requiert que le nombre d'observations par cluster soit grand. En présence de plusieurs clusters de petites tailles, il est plutôt approprié de substituer chaque observation par sa déviation par rapport à la moyenne du cluster, et d'effectuer une régression par moindres carrés ordinaires sur les données transformées¹¹ (Cameron et Trivedi, 2010). Ces méthodes imposent toutefois que les régresseurs qui ne varient pas à l'intérieur d'un même cluster soient retirés du modèle, ce qui peut être contraignant lorsque ces variables sont d'intérêt. Si tel est le cas, il est aussi possible d'omettre les effets fixes et d'intégrer au modèle un ensemble de régresseurs invariants au niveau du cluster. Cependant, pour corriger entièrement l'effet de la corrélation entre les observations, ces variables doivent capter la totalité de l'hétérogénéité entre les clusters, ce qui peut s'avérer être une condition difficile à satisfaire (Steenbergen et Jones, 2002)¹².

Le modèle de régression multiniveaux (connu également sous l'appellation de modèle mixte ou modèle hiérarchique), permet de relâcher cette hypothèse en intégrant un terme d'erreur aléatoire à chacun des niveaux d'analyse, captant l'hétérogénéité non-expliquée par les régresseurs. Il permet donc d'analyser les relations entre des variables de différents niveaux, la variable dépendante étant mesurée au niveau inférieur, tout en intégrant les dépendances entre les observations d'un même niveau (Hox, 1998 ; Hox, 2010). Cette approche est répandue en sciences sociales, politique et de la santé, pour étudier des effets contextuels (de niveau 2 ou supérieur) sur une variable

11. Voir Cameron et Trivedi (2010, p. 840) pour plus de détails sur ces méthodes.

12. Pour plus de détails sur les estimations avec données clusterisées, voir par exemple Wooldridge (2003) ou Cameron et Miller (2010).

individuelle (de niveau 1) (ex. Laird et Ware, 1982 ; Goldstein et Thomas, 1996 ; Steenbergen et Jones, 2002 ; Steele, Vignoles et Jenkins, 2007 ; Brumback, Zheng et Dailey, 2013). Bien que peu utilisées en économie jusqu'à présent, des analyses multiniveaux ont plus récemment été effectuées à partir de bases de données d'enquête ménages, notamment afin d'effectuer des analyses de pauvreté (Aassve et Arpino, 2007 ; Kim, Lee et Lee, 2010 ; Chen et Wang, 2015 ; Kim, Mohanty et Subramanian, 2016). La modélisation multiniveaux est appropriée pour cette analyse. En effet, la structure hiérarchique et clustérisée des données utilisées justifient la nécessité de tenir compte d'une potentielle corrélation entre les observations. De plus, certaines variables de niveau supérieur au ménage seront d'intérêt pour l'analyse et pourront ainsi être intégrées au modèle.

3.2 Le modèle multiniveaux

Le modèle multiniveaux peut être vu comme une généralisation du modèle de régression traditionnel dans lequel un ou des effets aléatoires autres que celui associé au terme d'erreur individuel sont inclus. Le modèle peut être défini sous la forme générale suivante :

$$\begin{aligned}
 Y &= X\beta + ZU + \epsilon \\
 \epsilon &\sim \mathcal{N}(0, \sigma_\epsilon^2)
 \end{aligned}
 \tag{1}$$

où la variable dépendante Y est un vecteur $n \times 1$, X est une matrice $n \times p$ de variables explicatives, β est un vecteur $p \times 1$ de paramètres inconnus, souvent appelés "effets fixes"¹³, et Z est une matrice $n \times q$ de variables explicatives associées à un vecteur U , $q \times 1$ d'effets aléatoires. Le vecteur ϵ , $n \times 1$, est l'erreur aléatoire individuelle, que l'on suppose normalement distribuée autour de 0, avec une variance constante. Dans ce modèle, l'effet fixe β est analogue à celui estimé par la méthode des moindres carrés ordinaires. Par ailleurs, l'effet aléatoire U n'est pas estimé directement ; il est plutôt caractérisé par sa variance. Ainsi, en faisant l'hypothèse que ϵ et U sont orthogonaux, la structure de la variance du modèle est définie telle que :

13. En modélisation multiniveaux, un effet fixe réfère à un coefficient qui ne varie pas entre les unités des niveaux supérieurs.

$$\text{Var} \begin{pmatrix} U \\ \epsilon \end{pmatrix} = \begin{pmatrix} \Sigma & 0 \\ 0 & \sigma_\epsilon^2 \end{pmatrix} \quad (2)$$

où Σ est la matrice de variance-covariance du vecteur d'effets aléatoires U .

Dans cet article, nous utilisons un modèle hiérarchique linéaire à deux niveaux. Les niveaux sont définis selon la structure de la base de données utilisée, les ménages constituant les unités du niveau 1, et les *wards* celles du niveau 2. Les VDC ne sont pas intégrés au modèle en tant que troisième niveau. En effet, bien qu'il n'y ait aucune règle concernant le nombre d'unités par niveau requis, des simulations ont montré qu'un nombre faible produisait des estimations biaisées. Aussi, un nombre d'unités supérieur à 20 est recommandé (Jia et al. 2011; Snijders et Bosker, 2012). Puisque l'échantillon ne contient que 10 VDC, un effet fixe contrôlant pour l'hétérogénéité entre les VDC sera plutôt ajouté au modèle. Le modèle linéaire à deux niveaux peut être représenté par une structure hiérarchique de modèles de régression dans laquelle un ou des coefficients associés aux variables explicatives du niveau inférieur ainsi que la constante sont des variables aléatoires définies au niveau supérieur. Soient l'indice i représentant l'unité du premier niveau et l'indice j , l'unité du second niveau. Alors, le premier niveau du modèle est défini par :

$$y_{ij} = \beta_{0j} + \sum_{p=1}^P \beta_{pj} x_{pij} + \epsilon_{ij} \quad (3)$$

où y_{ij} est la variable dépendante de l'observation i appartenant à l'unité j , x_{pij} sont les P variables explicatives de niveau 1 et $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ est le terme d'erreur individuel. Le second niveau du modèle comprend l'ensemble des équations caractérisant les paramètres aléatoires du premier niveau. Soient z_{qj} les Q variables explicatives de niveau 2. Ainsi, la constante est définie par :

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^Q \gamma_{0q} z_{qj} + u_{0j} \quad (4)$$

où γ_{00} est la constante moyenne, commune à toutes les unités du niveau 2 et u_{0j} est le terme aléatoire. Les coefficients associés aux variables explicatives du niveau inférieur sont quant à eux définis par :

$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^Q \gamma_{pq} z_{qj} + u_{pj}, \quad p = 1, \dots, P \quad (5)$$

où γ_{p0} est l'effet moyen la variable p sur la variable dépendante, commun à toutes les unités du niveau 2 et u_{pj} est le terme aléatoire. Mentionnons que le coefficient β_{pj} peut être supposé fixe¹⁴ et que les coefficients γ_{0q} et γ_{pq} peuvent être contraints à zéro de telle sorte que les coefficients aléatoires du premier niveau ne sont pas expliqués par l'ensemble des même régresseurs. En combinant (3), (4) et (5), on obtient le modèle à estimer :

$$y_{ij} = \underbrace{\left[\gamma_{00} + \sum_{p=1}^P \gamma_{p0} x_{pij} + \sum_{q=1}^Q \gamma_{0q} z_{qj} + \sum_{p=1}^P \sum_{q=1}^Q \gamma_{pq} z_{qj} x_{pij} \right]}_{X\beta} + \underbrace{\left[\sum_{p=1}^P u_{pj} x_{pij} + u_{0j} \right]}_{ZU} + \epsilon_{ij} \quad (6)$$

Le premier terme entre crochets est la partie fixe (analogue à $X\beta$ dans [1]) alors que le second terme entre crochets est la partie aléatoire (analogue à ZU dans [1]). Les effets aléatoires suivent une loi normale multivariée telle que :

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{00}^2 & \sigma_{01} \\ \sigma_{10} & \sigma_{11}^2 \end{pmatrix} \quad (7)$$

Ainsi u_{0j} et u_{1j} ont une variance constante et peuvent être corrélés entre eux. Par ailleurs, on suppose que les termes d'erreur entre les différents niveaux sont orthogonaux. Aussi, $Cov(u_{0j}, \epsilon_{ij}) = Cov(u_{1j}, \epsilon_{ij}) = 0$.

3.3 Les coefficients de régression intra-groupe et inter-groupe

Lorsque l'on estime un modèle multiniveaux, il est souvent utile de distinguer les coefficients de régression intra-groupe et inter-groupe associés à une variable explicative mesurée au niveau inférieur. En effet, si les observations appartenant à un même groupe (cluster) sont corrélées entre elles, une variable explicative mesurée au premier niveau peut affecter la variable dépendante à travers deux effets potentiellement distincts et indépendants : (1) un effet individuel ou intra-groupe

14. Alors $\beta_{pj} = \beta_p, \forall j$.

et (2) un effet de cluster ou inter-groupe. Ainsi, le coefficient de régression intra-groupe mesure l'effet individuel moyen de la variable explicative sur la variable dépendante, pour les observations appartenant à un même groupe alors que le coefficient inter-groupe capte l'effet de la moyenne intra-groupe de la variable explicative, sur la moyenne intra-groupe de la variable dépendante (Snijders et Bosker, 2012).

Différentes spécifications peuvent être utilisées afin de distinguer les coefficients intra-groupe et inter-groupe dans un modèle de régression¹⁵. Dans cet article, nous estimons le coefficient inter-groupe en intégrant au modèle la moyenne intra-groupe de la variable explicative, et le coefficient intra-groupe en substituant la variable explicative par sa valeur centrée par rapport à la moyenne intra-groupe. Formellement, supposons un modèle à deux niveaux, avec une seule variable explicative x_{ij} et sa moyenne intra-groupe $\bar{x}_{.j}$. Alors, pour obtenir les coefficients de régression intra-groupe et inter-groupe, le modèle estimé est :

$$y_{ij} = \gamma_{00} + \gamma_{10}(x_{ij} - \bar{x}_{.j}) + \gamma_{01}\bar{x}_{.j} + u_{0j} + \epsilon_{ij} \quad (8)$$

Ainsi, γ_{10} capte l'effet intra-groupe alors que γ_{01} capte l'effet inter-groupe. Mentionnons que (8) est statistiquement équivalent au modèle $y_{ij} = \tilde{\gamma}_{00} + \tilde{\gamma}_{10}x_{ij} + \tilde{\gamma}_{01}\bar{x}_{.j} + u_{0j} + \epsilon_{ij}$, dans lequel la variable x_{ij} n'est pas centrée par rapport à la moyenne intra-groupe. On observe en effet avec quelques manipulations que le coefficient inter-groupe $\gamma_{01} = \tilde{\gamma}_{10} + \tilde{\gamma}_{01}$ alors que le coefficient intra-groupe $\gamma_{10} = \tilde{\gamma}_{10}$. Cependant, Snijders et Bosker (2012) recommandent plutôt d'utiliser les variables centrées puisque l'estimation du coefficient inter-groupe est alors plus directe. De plus, les variables x_{ij} et $\bar{x}_{.j}$ sont susceptibles d'être corrélées entre elles. En utilisant la variable x_{ij} centrée par rapport à la moyenne intra-groupe, on élimine la colinéarité entre les variables (Angelson et al., 2014). En fait, les variables centrées deviennent orthogonales à toutes les variables du second niveau (Paccagnella, 2006). Ainsi, lorsque l'estimation des coefficients intra-groupe et intergroupes est d'intérêt, nous suivons donc la recommandation de Snijders et Bosker (2012).

15. Pour plus de détails, voir par exemple Snijders et Bosker (2012).

3.4 Méthode d'estimation

Trois principales méthodes sont généralement utilisées pour l'estimation d'un modèle multiniveaux, soient le maximum de vraisemblance, le maximum de vraisemblance restreint et l'estimation bayésienne (Snijders et Bosker, 2012). Cependant, lorsque les estimations sont effectuées à partir d'observations caractérisées par une probabilité de sélection inégale, l'intégration de poids d'échantillonnage est nécessaire pour éviter un biais. Alors, l'utilisation de la méthode du pseudo-maximum de vraisemblance est recommandée (Asparouhov, 2006 ; Rabe-Hesketh et Skrondal, 2006 ; Jia, Stokes et Wang, 2011) et sera donc employée dans cette analyse. L'utilisation de poids d'échantillonnage pour estimer un modèle de régression à un seul niveau est une procédure commune en microéconomie appliquée. Un vecteur de poids, proportionnel à l'inverse de la probabilité de sélection totale des observations, est alors introduit dans la méthode d'estimation. Si le plan d'échantillonnage comporte plusieurs phases, la probabilité de sélection totale est calculée en multipliant la probabilité de sélection des unités à chacune des phases.

Cependant, la théorie relative à l'incorporation des poids d'échantillonnage dans les analyses multiniveaux est encore peu développée. Des méthodes ont été proposées¹⁶ mais aucune ne fait l'unanimité (Asparouhov, 2006). De plus, il existe peu de travaux empiriques qui comparent les différentes méthodes, et peu de lignes directrices quant à la manière d'introduire les poids dans l'analyse (Carle, 2009). Dans la littérature, deux préoccupations liées à l'utilisation des poids d'échantillonnage dans les modèles multiniveaux ressortent. Premièrement, si les observations sont corrélées, intégrer un vecteur de poids proportionnel à l'inverse de leur probabilité de sélection totale, dans la fonction de vraisemblance, produira un biais (Pfeffermann et al., 1998). Aussi, plusieurs auteurs recommandent plutôt d'intégrer les vecteurs de poids associés aux unités de chaque niveau séparément (ex. Pfeffermann et al. 1998 ; Korn et Graubard, 2003 ; Asparouhov, 2006 ; Rabe-Hesketh et Skrondal, 2006). Dans cet article, nous utilisons la méthode d'estimation développée par Rabe-Hesketh et Skrondal (2006), qui adoptent cette approche pour l'intégration des poids d'échantillonnage dans la fonction de vraisemblance. La fonction de log-vraisemblance estimée est explicitée en annexe B.

Deuxièmement, des travaux de simulation montrent que la pondération des observations du premier

16. Voir par exemple Pfeffermann et al. 1998 ; Korn et Graubard, 2003 ; Kovacevic et Rai, 2003 ; Grilli et Pratesi, 2004 ; Rabe-Hesketh et Skrondal, 2006.

niveau génère une surestimation de la variance des effets aléatoires¹⁷ (ex. Pfeffermann, 1998 ; Rabe-Hesketh et Skrondal, 2006 ; Jia et al., 2011). Rabe-Hesketh et Skrondal (2006) l'expliquent analytiquement en montrant que les poids d'échantillonnage augmentent artificiellement la taille des clusters sans toutefois réduire la variance intra-groupe, ce qui a pour effet d'augmenter la variance totale. Cette hausse est alors incorrectement attribuée à la variance des effets aléatoires. Afin de réduire le biais, il est recommandé de transformer les poids associés aux unités du premier niveau, à l'aide d'une méthode de *scaling*, qui consiste à multiplier le vecteur de poids par un facteur d'échelle. Différents facteurs d'échelle ont été proposés et testés (ex. Pfeffermann, 1998 ; Asparouhov, 2006 ; Korn et Graubard, 2003 ; Rabe-Hesketh et Skrondal, 2006 ; Jia et al., 2011), mais il n'y a pas de consensus quant à celui produisant l'estimation la moins biaisée. Dans la littérature, deux principales méthodes applicables dans le cas d'un modèle à deux niveaux ont tout de même retenu l'attention, connues sous les noms de "méthode 1" et "méthode 2" de Pfeffermann et al. (1998)¹⁸. Soit $\omega_{i|j}$ le poids associé à l'unité i , conditionnel à la sélection de l'unité j , où i est l'unité du premier niveau et j l'unité du deuxième niveau.

1. Méthode 1 de Pfeffermann et al. (1998)

Le facteur d'échelle λ_j^1 est tel que la somme des poids attribués aux n_j observations appartenant à un cluster j , égale la taille "effective" du cluster¹⁹. Le poids ajusté par la "méthode 1", $\omega_{i|j}^{(1)}$, se calcule par :

$$\omega_{i|j}^{(1)} = \lambda_j^1 \omega_{i|j} = \omega_{i|j} \frac{\sum_{i=1}^{n_j} \omega_{i|j}}{\sum_{i=1}^{n_j} \omega_{i|j}^2} \quad (9)$$

2. Méthode 2 de Pfeffermann et al. (1998)

Le facteur d'échelle λ_j^2 est tel que la somme des poids attribués aux n_j observations appartenant à un cluster j , égale la taille de l'échantillon dans le cluster j , n_j . Le poids ajusté par la "méthode 2", $\omega_{i|j}^{(2)}$ se calcule par :

$$\omega_{i|j}^{(2)} = \lambda_j^2 \omega_{i|j} = \omega_{i|j} \frac{n_j}{\sum_{i=1}^{n_j} \omega_{i|j}} \quad (10)$$

17. Rappelons que les effets aléatoires sont les termes aléatoires associés aux équations des niveaux supérieurs au premier, tels que modélisés dans (4) et (5). Rappelons également que les effets aléatoires ne sont pas estimés directement, mais sont plutôt caractérisés par leur variance.

18. Pour la justification théorique de ces méthodes, voir Pfeffermann et al. (1998) ou Rabe-Hesketh et Skrondal (2006).

19. La taille effective est définie dans Potthoff et al. (1992).

Selon Snijders et Bosker (2012), les connaissances actuelles relatives à l'intégration des poids d'échantillonnage dans l'estimation multiniveaux ainsi qu'aux effets des différentes méthodes de *scaling* sur le biais ne sont pas suffisamment développées pour déterminer lequel des ces facteurs d'échelle devrait être utilisé. Les auteurs mentionnent également que la taille du biais peut dépendre du plan d'échantillonnage ainsi que du paramètre estimé (par exemple, le biais associé à la variance des effets aléatoires ne sera pas nécessairement égal à celui associé aux coefficients de régression). Dans la littérature, les travaux de simulation évaluant l'effet des poids d'échantillonnage et du *scaling* portent essentiellement sur le biais associé à l'estimation de la variance des effet aléatoires. Peu d'études, tant théoriques que de simulation, discutent des effets sur les coefficients de régression, donnant ainsi peu de lignes directrices pour l'estimation de ces paramètres. Or puisque dans cette analyse, nous nous intéressons principalement à l'impact des variables d'intérêts sur la variable dépendante, les résultats des simulations retrouvés dans la littérature ne suffisent pas pour guider nos choix méthodologiques. Dans ce contexte, nous suivons la recommandation de Carle (2009), de procéder à une simulation reproduisant le plan d'échantillonnage et la structure de la base de données utilisée, afin de vérifier l'effet de la pondération et de différentes méthodes de *scaling* sur les paramètres d'intérêt.

3.5 Simulation

L'objectif de la simulation est de vérifier l'effet de l'intégration des poids d'échantillonnage et de différents facteurs d'échelle dans la méthode d'estimation sur le biais des paramètres estimés, étant donné le plan d'échantillonnage et la structure de la base de données utilisée. Aussi, la paramétrisation doit permettre de reproduire le plus fidèlement possible les caractéristiques de la base et en particulier celles susceptibles de générer un biais. Pfeiffermann et al. (1998) soutiennent que les coefficients de régression seront convergent si le nombre de groupes est suffisamment grand, étant donnée la variance des effets aléatoires et des résidus. Par ailleurs, selon Jia et al. (2011), s'appuyant sur Pfeiffermann et al. (1998), Grilli et Pratesi (2004) et Asparouhov (2006), trois facteurs contribuent au biais de l'estimateur de la variance des effets aléatoires :

1. La taille des clusters : plus les clusters sont petits, plus le biais est important ;
2. La corrélation intra-classe²⁰ : plus la corrélation intra-classe est faible, plus le biais est

20. La corrélation intra-classe (ICC) mesure la dépendance entre les observations appartenant à un même cluster

important ;

3. L'informativité des poids²¹ : plus les poids sont informatifs, plus le biais est important.

Ainsi, en plus du plan d'échantillonnage, la paramétrisation doit permettre de reproduire la taille des clusters, la corrélation intra-classe ainsi que l'informativité des poids. La variable dépendante y est calculée telle que $y_{ij} = \beta_0 + \beta_{x_{ij}}x_{ij} + \beta_{x_j}x_j + u_j + e_{ij}$, où x_{ij} et x_j sont les variables explicatives de niveau ménage et *ward* respectivement, u_j est l'effet aléatoire et e_{ij} est le terme d'erreur. La paramétrisation est détaillée en annexe C.

Nous effectuons trois différentes simulations : (M0) avec poids d'échantillonnage, sans *scaling* ; (M1) avec poids d'échantillonnage et la méthode de *scaling* 1 ; (M2) avec poids d'échantillonnage et la méthode de *scaling* 2. La méthode d'estimation décrite en annexe B est utilisée pour estimer le modèle. Les résultats sont présentés dans le tableau 2.

TABLE 2 – Résultats des simulations

Estimation par pseudo-maximum de vraisemblance avec poids d'échantillonnage						
	M0		M1		M2	
Paramètre	Estimation	Biais rel.	Estimation	Biais rel.	Estimation	Biais rel.
β_0	1,0003	0,0003	1,0010	0,0010	1,0010	0,0010
$\beta_{x_{ij}}$	1,0003	0,0003	0,9996	-0,0003	0,9996	-0,0003
β_{x_j}	1,0006	0,0006	0,9993	-0,0007	0,9993	-0,0007
σ_u^2	0,1274	0,2745	0,0931	-0,0690	0,0931	-0,0690
σ_ϵ^2	0,3890	-0,0275	0,3994	-0,0015	0,3994	-0,0015

5 000 itérations. Vraies valeurs : $\beta_0 = 1$, $\beta_{x_{ij}} = 1$, $\beta_{x_j} = 1$, $\sigma_u^2 = 0,10$, $\sigma_\epsilon^2 = 0,40$. σ_u^2 est la variance de u . σ_ϵ^2 est la variance de ϵ .

Les deux méthodes de *scaling* (M1 et M2) génèrent des résultats identiques. De plus, les coefficients estimés ($\beta_{x_{ij}}$ et β_{x_j}), avec et sans *scaling*, convergent vers leur vraie valeur, le biais relatif étant très faible et variant entre -0,0007 et 0,0006. Le biais relatif calculé pour la variance de l'effet aléatoire σ_u^2 est plus important mais conforme à la littérature. En effet, l'estimation sans *scaling* surestime la variance de 27,45 % alors que celle avec *scaling*, la sous-estime de 6,90 %. L'estimation avec *scaling* génère ainsi un résultat plus fiable pour la variance des effets aléatoires. Un test de Wald confirme que le choix d'utiliser ou non un facteur d'échelle n'affecte pas significativement les coefficients de régression. Les résultats obtenus ne nous incitent donc pas à rejeter l'une ou l'autre des méthodes.

et sera définie à la section 5.

21. L'informativité des poids, telle que définie dans Pfeiffermann (1993), mesure la dépendance entre les poids d'échantillonnage et la variable dépendante.

Aussi, nous estimerons le modèle final et présenterons les résultats obtenus avec et sans *scaling*. Puisque les deux méthodes de *scaling* génèrent des résultats identiques, nous n'utiliserons que la méthode 1. Si le modèle est bien spécifié, les coefficients de régression estimés avec et sans *scaling* devraient être statistiquement égaux, ce que nous vérifierons par un test de Wald. Enfin, la variance des effets aléatoires devra être interprétée en tenant compte du biais estimé.

4 Les variables

Dans le tableau 3 se trouvent les moyennes pour chacune des variables du modèle, pour l'échantillon complet (colonne "Complet"), pour les ménages impliqués dans l'écotourisme (colonne "Impliqué") ainsi que pour les ménages qui ne sont pas impliqués dans l'écotourisme (colonne "Non-impliqué"). La dernière colonne contient les valeurs minimales et maximales des variables, pour l'échantillon complet.

4.1 La variable dépendante

Nous mesurons le bien-être des ménages par les dépenses annuelles de consommation. Dans la littérature, il est généralement reconnu que les dépenses sont plus lisses et moins affectées par des fluctuations de court terme que le revenu (Deaton et Zaidi, 2002). De plus, elles sont souvent considérées comme étant une meilleure mesure de bien-être monétaire puisqu'elles sont plus directement liées à la satisfaction des besoins de base et moins sujettes aux erreurs de mesure (Banque Mondiale, 2011). Les dépenses sont exprimées en équivalent adulte afin de tenir compte de l'hétérogénéité dans la composition des ménages ainsi que de la réallocation intra-ménage. Nous utilisons l'échelle d'équivalence d'Oxford, qui accorde un poids d'une unité de consommation au premier adulte, 0,7 unité aux autres membres du ménage âgés de 14 ans et plus et 0,5 unité aux enfants de moins de 14 ans. Le vecteur de consommation est calculé selon la procédure recommandée par Deaton et Zaidi (2002), utilisée dans le cadre des enquêtes nationales sur les ménages du Nepal Living Standard Survey (NLSS). La dépense de consommation moyenne exprimée par équivalent adulte pour l'ensemble des ménages de l'échantillon est de 160 257 NPR. De plus, on observe un écart de bien-être important et significatif entre les ménages impliqués dans l'écotourisme et ceux qui ne le sont pas. En effet, la consommation moyenne pour le premier groupe est de 264 782 NPR,

ce qui est près du double de celle du second groupe.

4.2 Les variables d'intérêt

Cinq variables sont construites afin de mesurer l'effet de l'écotourisme et des restrictions environnementales sur le bien-être local.

4.2.1 Écotourisme

Tel que soulevé par Nepal (2002), il existe plusieurs interprétations quant au concept de l'écotourisme et distinguer les activités écotouristiques des autres peut être ambigu. Par ailleurs, Eagles (1992)²² définit les écotouristes comme ceux participant à des expériences orientées vers la nature dans des environnements naturels vierges ou sauvages, ce qui correspond aux activités proposées dans les zones protégées du Népal²³. Aussi, nous considérons que toutes les activités touristiques se déroulant à l'intérieur des zones protégées de notre échantillon appartiennent à la catégorie de l'écotourisme. Nous utilisons d'abord la question d'enquête : "Est-ce que vous ou un membre de votre ménage avez tiré un bénéfice économique provenant de touristes durant les 12 derniers mois ?" (réponse : oui/non), afin de distinguer les ménages que nous qualifierons "d'impliqués dans l'écotourisme", de ceux "non-impliqués dans l'écotourisme"²⁴. Ensuite, à partir des ménages impliqués dans l'écotourisme, nous créons deux catégories : (1) les ménages impliqués à titre de salariés et (2) les ménages impliqués à titre de travailleurs autonomes. Ainsi, un ménage appartient à la première catégorie si au moins un individu dans le ménage est impliqué dans l'écotourisme en tant que salarié, et à la seconde si au moins un individu dans le ménage est impliqué dans l'écotourisme en tant que travailleur autonome. De plus, un ménage peut appartenir aux deux catégories, si au moins un individu dans le ménage est salarié et au moins un individu est travailler autonome. Nous distinguons ces deux catégories d'emplois puisqu'un écart de bien-être important semble exister entre les ménages impliqués comme salariés, et ceux impliqués comme travailleurs autonomes. En effet, la consommation moyenne des ménages appartenant à la première catégorie est de 131 288

22. Cité dans Holden et Sparrowhawk (2002).

23. Voir welcomenepal.com pour plus de détails.

24. A noter qu'il ne s'agit pas d'un indicateur de revenu. L'objectif de cette question est essentiellement de vérifier si, parmi les différents secteurs d'activité générateurs de revenus, le tourisme en est un dans lequel un ou des membres du ménage sont impliqués.

roupies népalaises (NPR) alors que celle des ménages appartenant à la deuxième catégorie est de 346 279 NPR. Ainsi, deux indicateurs sont créés, avec les exposants *sal*=salarié et *aut*=travailleur autonome :

1. ind^{sal} : L'indicateur prend la valeur de 1 si le ménage est impliqué dans l'écotourisme à titre de salarié et de 0 sinon ;
2. ind^{aut} : L'indicateur prend la valeur de 1 si le ménage est impliqué dans l'écotourisme à titre de travailleur autonome et de 0 sinon.

Nous transformons ensuite ces indicateurs afin de construire des variables permettant de distinguer les coefficients intra-groupe et inter-groupe, selon la procédure décrite à la section 3.3. Ainsi, nous calculons d'abord la moyenne des indicateurs ind^{sal} et ind^{aut} , pour chacun des *wards*. Ces variables, nommées respectivement $tourward^{sal}$ et $tourward^{aut}$, permettront d'obtenir les coefficients inter-groupe. Intuitivement, elles mesurent la part des ménages par *ward* impliqués dans l'écotourisme, comme salariés ou travailleurs autonomes. Puis, afin d'obtenir les variables associées aux coefficients intra-groupes, nous soustrayons aux indicateurs ind^{sal} et ind^{aut} , leur moyenne intra-groupe respective. Ainsi, les variables centrées par rapport à leur moyenne intra-groupe, nommées $tour^{sal}$ et $tour^{aut}$, captent l'effet pour un ménage d'être impliqué dans l'écotourisme - en tant que salarié ou travailleur autonome. Les variables intégrées au modèle de régression sont donc ainsi définies :

1. $tourward^{sal}$: Part des ménages du *ward* impliqués dans l'écotourisme comme salariés (moyenne intra-groupe de ind^{sal})
2. $tour^{sal}$: Indicateur ind^{sal} centré par rapport à la moyenne du *ward* ($ind^{sal} - tourward^{sal}$).
3. $tourward^{aut}$: Part des ménages du *ward* impliqués dans l'écotourisme comme travailleurs autonomes (moyenne intra-groupe de ind^{aut})
4. $tour^{aut}$: Indicateur ind^{aut} centré par rapport à la moyenne du *ward* ($ind^{aut} - tourward^{aut}$)

4.2.2 Restrictions environnementales

Pour mesurer l'effet d'être contraint au niveau de l'utilisation des ressources naturelles, sur le bien-être local, nous construisons une variable à partir de deux questions de l'enquête. Les ménages interrogés sont tous résidents d'une zone protégée et soumis à une certaine réglementation plus

ou moins stricte relative à l'utilisation des ressources naturelles. Certains ménages affirment que si cette réglementation n'était pas en place, ils utiliseraient davantage de ressources naturelles (ces ménages sont ainsi référencés comme "contraints") alors que d'autres soutiennent que même sans la réglementation, ils n'utiliseraient pas plus de ressources (ménages "non-contraints"). Aussi, pour distinguer les ménages "contraints" des ménages "non-contraints", nous utilisons les deux questions d'enquête suivantes :

1. Si votre village n'appartenait pas à une zone protégée, vous collecteriez plus de ressources dans la forêt. Êtes-vous : d'accord / pas d'accord / ne savez pas.
2. Si votre village n'appartenait pas à une zone protégée, votre production agricole serait plus importante. Êtes-vous : d'accord / pas d'accord / ne savez pas.

Nous interprétons que les ménages étant d'accord avec l'une de ces affirmations sont contraints dans leur utilisation des ressources naturelles, compte tenu des restrictions imposées par la mise en place de la zone protégée. Nous créons une variable de catégorie prenant la valeur de 2 si le ménage affirme être d'accord avec les deux affirmations, de 1 s'il est d'accord avec seulement l'une des affirmations, et de 0 sinon. Ainsi, plus le ménage se dit affecté par les restrictions, plus la valeur de la variable est élevée. La variable intégrée au modèle est donc la suivante :

1. *cont* : Indicateur de contrainte dans l'utilisation des ressources naturelles. Prend la valeur de 0, 1 ou 2, 0 indiquant que le ménage ne considère pas être contraint et 2 qu'il considère être très contraint.

A noter que nous n'estimons pas les coefficients de régression inter-groupe et intra-groupe associés à cette variable. En effet, nous avons testé plusieurs spécifications au modèle en intégrant la moyenne intra-groupe et dans chacun des cas, l'effet de la variable sur le bien-être était non-significatif. Intuitivement, ce résultat n'est pas surprenant puisque la variable de restriction environnementale est construite à partir de questions subjectives individuelles ; il est donc peu probable que la part des ménages dans un *ward* se disant contraints affecte directement le bien-être moyen dans le *ward*. La moyenne intra-groupe de la variable de restriction environnementale est donc omise afin de ne pas alourdir le modèle.

4.2.3 Test de stabilité structurelle

Puisque les opportunités économiques générées par la mise en place d'une zone protégée sont susceptibles d'inciter les ménages à y migrer, un potentiel biais de sélection ne peut être exclu. Aussi, nous testons la stabilité structurelle entre les ménages résidant sur le territoire protégé avant la désignation de la zone (49,5 % de l'échantillon), et ceux ayant migré vers la zone après sa mise en place (50,5 % de l'échantillon) à l'aide d'un test de Chow. Tous les coefficients associés aux variables d'intérêt sont statistiquement égaux entre les deux sous-échantillons, à l'exception de ceux pour la variable *tour^{sal}* qui par ailleurs est fortement non-significative. Nous concluons que la migration des ménages vers les zones protégées ne biaise pas les résultats.

4.3 Les variables de contrôle

La sélection des variables de contrôle ajoutées au modèle est basée sur une revue de la littérature sur la relation entre zones protégées, tourisme et bien-être (Ferraro et Hanauer, 2010 ; Sims, 2010 ; Ferraro, Hanauer et Sims, 2011 ; Richardson et al. 2012 ; Canavire-Bacarreza et Hanauer, 2013), sur la pauvreté au Népal (Bhatta et Sharma, 2006 ; Baland et al., 2010 ; Lokshin et al. 2010) ainsi que sur la disponibilité des données. Comme caractéristique du ménage, nous incluons le sexe et l'âge du chef de ménage, la taille du ménage ainsi que la part des individus dans le ménage âgés de moins de 15 ans et de plus de 60 ans. De plus, une variable indique si le ménage appartient à une caste supérieure (Brahmin, Chhetri ou Newar) et une autre indique si le ménage appartient à la caste inférieure des Dalit. Nous mesurons le capital humain par le niveau d'éducation maximal atteint dans le ménage. Pour tenir compte de l'emploi et des autres sources de revenu, nous incluons la part des individus dans le ménage : (1) âgés de 15 ans et plus et sans emploi, (2) travaillant dans le secteur agricole et (3) travaillant de manière autonome dans un secteur excluant l'agriculture et le tourisme. Nous ajoutons également une variable indiquant si le ménage a reçu des transferts internationaux durant les 12 derniers mois. Pour tenir compte du capital physique détenu par le ménage, nous incluons la taille de la terre possédée ainsi que la valeur de la résidence. L'accès aux infrastructures est considéré par l'ajout de variables relatives à la distance par rapport au marché, à un arrêt d'autobus²⁵, à une école primaire et secondaire ainsi qu'à un centre de santé. Enfin,

25. Dans la majorité de villages enquêtés, il n'y avait pas d'arrêt d'autobus formel. Les individus devaient se rendre le long de la route principale et signaler au conducteur de l'autobus de s'arrêter lorsqu'il passait. Cette

une variable indique si le ménage a collecté du bois de combustion dans la forêt au cours des 12 derniers mois.

TABLE 3 – Moyennes des variables pour l'échantillon complet et sous-échantillons

Description des variables	Complet	Impliqué	Non-impliqué	Min-max
Variable dépendante				
Dépenses de consommation (en NPR)***	160 257	264 782	138 979	11 259-8 440 485
Variables d'intérêt				
Restriction environnementale	0,30	0,35	0,29	0-2
Implication dans tourisme - salarié \pm	0,07	0,40	0,00	0-1
Implication dans le tourisme - tr. autonome \pm	0,12	0,68	0,00	0-1
Part du <i>ward</i> dans tourisme - salarié***	0,07	0,18	0,05	0-0,5
Part du <i>ward</i> dans tourisme - tr. autonome***	0,12	0,29	0,08	0-1
Variable de contrôle				
Sexe du chef de ménage (=1 si masculin)**	0,78	0,84	0,77	0-1
Age du chef de ménage	47,03	45,48	47,34	16-96
Taille du ménage*	4,01	4,27	3,96	1-17
Part du ménage âgé <15 et >60 ans***	0,34	0,27	0,35	0-1
Caste supérieure (=1 si oui)***	0,33	0,24	0,35	0-1
Dalit (=1 si oui)***	0,06	0,03	0,07	0-1
Niveau d'éducation max. du ménage***	10,54	11,67	10,31	0-21
Part du ménage sans emploi***	0,30	0,21	0,32	0-1
Part du ménage dans le secteur agricole***	0,45	0,32	0,47	0-1
Part du ménage étant tr. autonomes \mp	0,12	0,10	0,12	0-1
Transferts internationaux reçus (=1 si oui)***	0,26	0,17	0,28	0-1
Taille de la terre (en ropani)***	8,36	5,51	8,94	0-160,74
Log. de la valeur de la résidence (en NPR)***	13,65	14,20	13,55	9,90-18,42
Distance au marché \diamond	2,80	2,73	2,81	1-5 \diamond
Nombre d'institutions à moins de 2 km \diamond	2,28	2,25	2,29	0-3
Arrêt d'autobus à moins de 0,5 km (=1 si oui)	0,41	0,42	0,41	0-1
Collecte de bois dans la forêt (=1 si oui)	0,56	0,52	0,56	0-1
Nombre d'observations	1563	275	1 288	

\pm Variable non-déviée de la moyenne du groupe, prenant la valeur de 1 lorsque le ménage est impliqué et 0 sinon.
 \mp Dans un secteur autre que agricole et touristique. \diamond =1 : plus de 10 km, =2 : entre 5 et 10 km, =3 : entre 2 et 5 km, =4 : entre 0,5 et 2 km, =5 : moins de 0,5 km. \diamond Parmi école primaire, école secondaire et centre de santé. ***Significativité à 1 % de l'écart entre les moyennes des sous-échantillons "impliqué" et "non-impliqué". **Significativité à 5 %. *Significativité à 10 %.

5 Résultats

Les résultats obtenus pour l'estimation de deux différentes spécifications du modèle sont présentés. Dans la section 5.1, les résultats de l'estimation d'un modèle vide à constante aléatoire permettent de confirmer la présence d'hétérogénéité entre les *wards*. Dans la section 5.2, un modèle à coefficient variable capte donc également la distance par rapport une route principale.

aléatoire est spécifié et estimé pour mesurer l'effet de l'écotourisme et des restrictions environnementales sur le bien-être des ménages. Les résultats obtenus pour l'estimation sans et avec les variables de contrôle sont présentés.

5.1 Modèle vide à constante aléatoire

Le premier modèle estimé est un modèle "vide" dans lequel aucune variable explicative n'est introduite, seule une constante est aléatoire. Cette spécification est utile pour confirmer la pertinence de l'utilisation du modèle multiniveaux puisqu'elle permet de déterminer s'il y a bien présence d'hétérogénéité entre les groupes. Le modèle estimé, combinant les niveaux 1 et 2 est :

$$y_{ij} = \gamma_{00} + u_j + \epsilon_{ij} \quad (11)$$

où y_{ij} est le bien-être du ménage i dans le *ward* j , γ_{00} est une constante, u_j est un effet aléatoire, invariant dans le *ward* j et ϵ_{ij} est un terme d'erreur individuel. Dans ce modèle, γ_{00} s'interprète comme le bien-être moyen dans la population et u_j comme les caractéristiques affectant le bien-être, communes aux ménages appartenant à un même *ward* et expliquant l'hétérogénéité entre les groupes. Ainsi, $(\gamma_{00} + u_j)$ est une mesure du bien-être moyen dans le *ward* j alors que le bien-être de chaque ménage $i \in j$ dévie de cette moyenne par ϵ_{ij} (Snijders et Bosker, 2012). Les *wards* j varient aléatoirement les unes des autres avec une variance inter-groupe, σ_u^2 , et les ménages i varient à l'intérieur du *ward* j avec une variance intra-groupe, σ_ϵ^2 . Les résultats sans et avec *scaling* sont présentés dans le tableau 4.

TABLE 4 – Résultats de l'estimation du modèle à constante aléatoire

	Dépenses de consommation exprimées en équivalent adulte (logarithme)	
	Sans scaling	Avec scaling
Effet fixe		
Constante	11,564*** (0,122)	11,567*** (0,127)
Effets aléatoires		
σ_u^2	0,157 (0,028)	0,133 (0,027)
σ_ϵ^2	0,419 (0,023)	0,466 (0,052)
ICC	0,273	0,222
Observations	1 563	1 563

Écart-types robustes entre parenthèses. ***Significativité à 1 %

La variance inter-groupe estimée est de 0,157 sans *scaling* et de 0,133 avec *scaling*. Aussi, conformément aux résultats de la simulation, l'estimation sans *scaling* est supérieure à celle avec *scaling*. La variance intra-groupe estimée, est de 0,419 sans *scaling* et de 0,466 avec *scaling*. A partir de ces résultats, nous calculons la corrélation intra-classe (ICC) qui permet de vérifier la présence d'hétérogénéité entre les groupes. L'ICC mesure l'importance de la dépendance entre les observations appartenant à un même groupe et est utilisée pour confirmer la pertinence de l'utilisation d'un modèle multiniveaux, relativement à un modèle linéaire à un seul niveau. Plus l'ICC est élevé, plus les observations appartenant à un même groupe sont corrélées, et plus l'hétérogénéité entre les groupes est importante. Formellement, puisque les termes d'erreur sont supposés indépendants, la variance totale σ^2 se décompose telle que $\sigma^2 = \sigma_u^2 + \sigma_\epsilon^2$ et la corrélation intra-classe se calcule par :

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}$$

L'ICC calculée à partir des estimations sans *scaling* est de 0,273, et celle calculée à partir des estimations avec *scaling* est de 0,222. Ainsi, près du quart de la variance totale de la variable dépendante s'explique par des caractéristiques communes aux ménages appartenant à un même *ward*. Un test de Fisher vérifie que dans les deux cas, l'ICC est significativement différente de 0. Ce résultat confirme l'existence d'un écart de bien-être significatif entre les *wards* ; l'utilisation d'un modèle multiniveaux est donc appropriée pour cette analyse.

5.2 Modèle à coefficients aléatoires

Pour mesurer l'effet de l'écotourisme et des restrictions environnementales sur le bien-être des ménages, nous estimons le modèle :

$$y_{ij} = \beta_{0j} + \beta_{1j}tour_{ij}^{sal} + \beta_{2j}tour_{ij}^{aut} + \beta_3cont_{ij} + \sum \beta_p X_{pij} + \epsilon_{ij} \quad (12)$$

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}tourward_j^{sal} + \gamma_{02}tourward_j^{aut} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} + u_{2j} \end{aligned} \quad (13)$$

où (12) est la spécification du premier niveau du modèle et (13) celle le second niveau. La constante β_{0j} ainsi que les coefficients β_{1j} et β_{2j} sont aléatoires. La constante β_{0j} , interprétée comme le niveau de bien-être moyen par *ward*, étant donnés les régresseurs dans (12), est expliquée par la part des ménages dans le *ward* impliqués dans l'écotourisme, comme salariés ($tourward_j^{sal}$) et travailleurs autonomes ($tourward_j^{aut}$). La variabilité de la constante qui n'est pas expliquée par ces deux variables est attribuée au terme aléatoire u_{0j} . Les coefficients β_{1j} et β_{2j} , quant à eux, varient uniquement en fonction du terme aléatoire. Ainsi, on suppose que l'effet de l'écotourisme sur le bien-être varie aléatoirement entre les *wards*. Mentionnons également que le modèle multiniveaux permet d'estimer la covariance entre les différents termes aléatoires d'un même niveau. Cependant, dans ce modèle, toutes les covariances estimées sont non-significativement différentes de 0. Aussi, elles ont été contraintes à 0 afin de ne pas alourdir l'estimation. Le coefficient associé à la variable de restrictions environnementale, β_3 , est fixe. En effet, associer un terme aléatoire à ce coefficient n'améliore pas la spécification du modèle. De plus, sa variance a été testée non-significative, indiquant que l'effet de la variable de restrictions environnementale sur le bien-être ne varie pas entre les *wards*. Enfin, X_{pij} est un vecteur de variables de contrôle. Tous les coefficients associés à ces variables sont fixes. Les résultats de l'estimation du modèle, sans les variables de contrôle (Modèle 1) et avec les variables de contrôle (Modèle 2), sans et avec *scaling*, sont présentés dans le tableau 5.

Dans la section 3.5, nous avons testé et validé l'égalité statistique entre les coefficients de régression simulés sans et avec *scaling*. Aussi, si (12) et (13) sont bien spécifiées, les coefficients estimés sans et avec *scaling* devraient être statistiquement égaux. Un test de Wald confirme l'égalité statistique pour les variables d'intérêt²⁶. En ce qui concerne la variance des effets aléatoires, nous observons que l'estimation sans *scaling* est plus élevée que celle avec *scaling*, ce qui est également conforme aux résultats de la simulation et de la littérature.

Les résultats obtenus montrent d'abord que, tel qu'anticipé, l'effet de l'écotourisme sur le bien-être des ménages diffère selon la catégorie d'emploi. En effet, l'implication dans l'écotourisme pour un ménage, en tant que salarié ($tour^{sal}$) de même que la part des ménages par *ward* impliqués dans l'écotourisme en tant que salariés ($tourward^{sal}$) n'affectent pas significativement le bien-être. Cependant, l'implication dans l'écotourisme pour un ménage, en tant que travailleur autonome

26. La statistique calculée suit une loi de Student avec 1 562 degrés de liberté et la covariance entre les coefficients est obtenue par simulation avec 500 itérations.

TABLE 5 – Résultats de l'estimation du modèle à coefficients aléatoires

	Dépenses de consommation exprimées en équivalent adulte (logarithme)			
	Modèle 1		Modèle 2	
	Sans scaling	Avec scaling	Sans scaling	Avec scaling
Effets fixes				
Constante	11,387*** (0,136)	11,388*** (0,142)	11,019*** (0,219)	11,081*** (0,164)
Tourisme individuel - salarié	0,061 (0,081)	0,075 (0,079)	0,124 (0,081)	0,101 (0,074)
Tourisme ind. - autonome	0,669*** (0,125)	0,735*** (0,118)	0,500*** (0,103)	0,517*** (0,093)
Restriction environnementale	-0,022 (0,024)	-0,045 (0,020)	0,018 (0,026)	-0,018 (0,020)
Tourisme <i>ward</i> - salarié	-0,558 (0,406)	-0,660* (0,398)	-0,374 (0,515)	-0,340 (0,446)
Tourisme <i>ward</i> - autonome	1,172*** (0,331)	1,308*** (0,312)	1,034*** (0,381)	1,084*** (0,421)
Contrôles	non	non	oui	oui
Effets aléatoires				
σ_u^2	0,010 (0,025)	0,073 (0,028)	0,059 (0,009)	0,034 (0,009)
$\sigma_{tour^{si}}^2$	0,091 (0,036)	0,000 (0,000)	0,103 (0,038)	0,000 (0,000)
$\sigma_{tour^{ai}}^2$	0,200 (0,055)	0,071 (0,046)	0,189 (0,061)	0,069 (0,042)
σ_ϵ^2	0,360 (0,017)	0,392 (0,022)	0,260 (0,020)	0,280 (0,020)
Observations	1 563	1 563	1 563	1 563

***Significativité à 1 %. **Significativité à 5 %. *Significativité à 10 %. Écart-types robustes entre parenthèses. Effet fixe VDC ajouté au modèle. Covariances entre les effets aléatoires contraintes à 0.

($tour^{aut}$) affecte positivement et significativement (à 1 %) le bien-être. De plus, l'effet est considérable : s'impliquer dans une occupation directement liée à l'écotourisme en tant que travailleur autonome générerait une hausse des dépenses de consommation de près de 65 %²⁷. On observe également que le coefficient associé à la part des ménages dans le *ward*, impliqués dans l'écotourisme en tant que travailleurs autonomes ($tourward^{aut}$) est positif et significatif. De plus, sa valeur supérieure à celle du coefficient associé à la variable $tour^{aut}$ indique que ce secteur d'activité produit une externalité positive sur le bien-être. En effet, tel que discuté dans la section 3.3, le coefficient inter-groupe peut s'exprimer en une somme de deux termes. Premièrement, le coefficient intra-groupe, c'est-à-dire celui associé à la variable $tour^{aut}$, et deuxièmement, un effet additionnel qui se calcule par $tourward^{aut} - tour^{aut}$. Intuitivement, le coefficient inter-groupe mesure l'effet de l'implication des ménages du *ward* dans l'écotourisme en tant que travailleurs autonomes, sur le bien-être moyen dans le *ward* et constitue la somme de deux effets : (1) celui de l'implication des ménages dans l'écotourisme, mesuré par le coefficient associé à la variable $tour^{aut}$ et (2) une

27. $\Delta\%y = 100(\exp[\beta] - 1)$.

contribution additionnelle, que nous interprétons comme une externalité produite par le secteur, sur le bien-être des ménages appartenant au *ward*. Autrement dit, le premier effet capte la hausse du bien-être moyen dans le *ward*, générée par la hausse du bien-être des ménages s’impliquant dans l’écotourisme et le deuxième effet capte une externalité profitant à tous les ménages appartenant au *ward*, indépendamment de leur implication individuelle. Ainsi, une hausse de la part des ménages du *ward* impliqués dans l’écotourisme en tant que travailleurs autonomes de 10 points de pourcentage aurait un effet multiplicatif total de 1,1089 (incluant les effets [1] et [2], sans *scaling*) sur la moyenne des dépenses de consommation dans le *ward*²⁸. L’externalité (effet [2]), quant à elle, a un effet multiplicatif de 1,0549.

La variable *cont*, intégrée au modèle afin de mesurer l’effet pour un ménage d’être contraint au niveau de l’utilisation des ressources naturelles, est non-significative. Ce résultat suggère que les populations locales dans les zones protégées ont pu s’adapter aux restrictions environnementales imposées. En effet, les zones protégées visitées ont été établies entre 1973 et 1992. Les ménages résidant à l’intérieur de ces zones, qu’ils se disent contraints ou non, sont ainsi susceptibles d’avoir développé des mécanismes leur permettant de maintenir un certain niveau de consommation, malgré les restrictions. Cependant, ces mécanismes ne semblent pas être optimaux pour l’ensemble des ménages puisque ceux affirmant être contraints considèrent que leur bien-être serait supérieur s’ils avaient la possibilité d’utiliser davantage de ressources.

La variance des effets aléatoires diffère considérablement entre les estimations sans et avec *scaling*, bien qu’elle soit toujours conforme aux résultats de la simulation. En effet, la variance du coefficient associé à la variable d’implication individuelle dans l’écotourisme comme salarié est de 0,103 sans *scaling* alors qu’elle tend vers 0 avec *scaling*. La variance du coefficient associé à la variable d’implication individuelle dans l’écotourisme comme travailleur autonome, quant à elle, est de 0,189 sans *scaling* et de 0,069 avec *scaling*. Afin de vérifier que la variance des effets aléatoires est significativement différente de 0, nous procédons à un test du ratio de vraisemblance²⁹. Le test est effectué pour chacun des deux coefficients aléatoires, pour les estimations sans et avec *scaling*. Premièrement, la variance du coefficient associé à la variable *tour^{sal}* est significative à 1

28. $\Delta y = \exp(0,1 \times \beta)$.

29. La distribution de la statistique du ratio de vraisemblance n’est pas connue lorsque l’estimation est réalisée par pseudo-maximum de vraisemblance. Cependant, il s’agit à notre connaissance du seul test disponible pour vérifier la significativité de la variance des coefficients. Dans ce contexte, nous supposons que la statistique suit une loi du chi-carré et interprétons les résultats avec précaution.

% sans *scaling* et non-significative avec *scaling*. Cependant, puisque la variance sans *scaling* est surestimée alors que celle avec *scaling* est sous-estimée, il n'est pas possible de conclure sur la significativité de cette variance. Par ailleurs, la variance du coefficient associé à la variable *tour^{aut}* est significative à 1 % sans et avec *scaling*. Ainsi, l'effet pour un ménage d'être impliqué dans l'écotourisme en tant que travailleur autonome semble varier entre les différents *wards*. On estime que pour 95 % des *wards*, la variation des dépenses de consommation d'un ménage, suite à son implication dans l'écotourisme en tant que travailleur autonome est comprise entre -29,53 % et 285,74 % sans *scaling*, et entre 0,00 % et 182,92 % avec *scaling*³⁰. Ainsi, il semble que de s'impliquer dans l'écotourisme en tant que travailleur autonome est très peu susceptible d'affecter le bien-être négativement. Ce résultat est préliminaire mais suggère que certaines caractéristiques de la population communes aux ménages appartenant à un même *ward*, ou certaines caractéristiques du *ward* lui-même, affectent l'effet de l'écotourisme sur le bien-être. Identifier ces caractéristiques dans une recherche future permettrait de préciser les conditions permettant d'optimiser l'effet du développement écotouristique sur le bien-être.

6 Test de la distance de Cook

6.1 Robustesse de l'effet de groupes

Afin de s'assurer que les résultats obtenus ne sont pas biaisés par l'effet de l'un ou de quelques *wards* trop influents, nous procédons à un test de la distance de Cook, tel que proposé dans Snijders et Berkhof (2008)³¹. Ce test est analogue à celui de la distance de Cook pouvant être utilisé après une régression par moindres carrés pour vérifier la présence de valeurs aberrantes (Snijders et Berkhof, 2008). En analyse multiniveaux, le test de la distance de Cook mesure l'influence d'une unité du deuxième niveau sur la valeur des paramètres estimés (Möhring et Schmidt, 2012) et consiste à estimer le modèle en éliminant les unités du deuxième niveau, une à la fois. Soient p le nombre de paramètres fixes, $\hat{\beta}$ le vecteur de coefficients fixes estimés et $\hat{\Sigma}_F$ l'estimation de la matrice de covariance des paramètres fixes. L'indice $-j$ indique que le *ward* j est éliminé de l'estimation. Pour

30. Environ 95 % de la probabilité d'une distribution normale est comprise entre 1,96 écart-types de la moyenne. Ainsi, environ 95 % des *wards* ont un coefficient compris entre $\beta_{tour^{aut}} - 1,96 \times \sqrt{\sigma_{tour^{aut}}^2}$ et $\beta_{tour^{aut}} + 1,96 \times \sqrt{\sigma_{tour^{aut}}^2}$.

31. Pour plus de détails sur ce test, voir Snijders et Bosker (2012, p. 167-172)

mesurer l'influence du *ward* j sur les paramètre fixes, on calcule :

$$C_j^F = \frac{1}{p}(\hat{\beta} - \hat{\beta}_{-j})^T \hat{\Sigma}_{F(-j)}^{-1}(\hat{\beta} - \hat{\beta}_{-j}) \quad (14)$$

De même, la mesure de l'influence du *ward* j sur les paramètre aléatoires se calcule par :

$$C_j^A = \frac{1}{q}(\hat{U} - \hat{U}_{-j})^T \hat{\Sigma}_{A(-j)}^{-1}(\hat{U} - \hat{U}_{-j}) \quad (15)$$

avec \hat{U} , le vecteur de paramètres aléatoires et q le nombre de paramètres aléatoires estimés. On obtient la mesure de l'influence totale du *ward* j sur l'ensemble des paramètres estimés en calculant :

$$C_j = \frac{1}{p+q}(pC_j^F + qC_j^A) \quad (16)$$

Les mesures C_j^F et C_j^A sont comparées à un seuil critique, proposé par Belsley et al. (1980), égal à $\frac{4}{J}$, où J est le nombre total de *wards*. Les *wards* pour lesquels l'une de ces mesures est supérieure au seuil sont considérés comme ayant une influence importante sur les estimations. Snijders et Bosker (2012) expliquent que si le modèle est bien spécifié et que les variables explicatives sont approximativement distribuées de manière aléatoire entre les groupes, alors la mesure de l'influence d'un groupe devrait être à peu près proportionnelle à sa taille. Ainsi, un *ward* ayant une mesure supérieure au seuil critique, mais non-proportionnelle à sa taille, est susceptible de biaiser les estimations. Les résultats du test de la distance Cook, effectué à partir des estimations du modèle sans *scaling*, sont présentés en annexe D. Parmi les 71 *wards* de l'échantillon, 31 ont une mesure supérieure au seuil de Belsley et al. Afin d'identifier les *wards* associés à une mesure influente mais non-proportionnelle à leur taille, nous calculons le ratio entre la mesure de Cook et la taille du *ward*, pour les 31 *wards* ayant une mesure supérieure au seuil. Un ratio constant indiquerait une mesure proportionnelle à la taille, pour la totalité des *wards* influents. Le ratio calculé a une valeur moyenne de 0,00009 et un écart-type de 0,0001583. Nous observons que 4 des 31 *wards* influents ont un ratio dont la valeur est supérieure à la moyenne de plus d'un écart-type. Afin de vérifier leur influence sur les variables d'intérêt, nous estimons le modèle en éliminant ces quatre *wards*, un à la fois. Le tableau 6 contient les coefficients estimés des variables d'intérêt, pour ces 4 estimations. Ainsi, M0 réfère au modèle estimé à partir de l'échantillon complet et les modèles M1-M4 réfèrent

TABLE 6 – Comparaison des coefficients pour les variables d'intérêt, pour les estimations éliminant les *wards* influents

	M0	M1	M2	M3	M4
Tourisme individuel - salarié	-0,012	-0,013	-0,048	-0,012	-0,011
Tourisme individuel - autonome	0,524***	0,553***	0,520***	0,523***	0,526***
Restriction environnementale	-0,002	-0,002	-0,002	-0,001	-0,003
Tourisme <i>ward</i> - salarié	-0,367	-0,174	0,060	-0,505	-0,554
Tourisme <i>ward</i> - autonome	1,033**	0,784**	1,070**	1,244**	1,367***

***Significativité à 1 %. **Significativité à 5 %.

aux modèles estimés en éliminant l'un des *wards* influents.

Nous observons que le fait d'éliminer un *ward* influent ne modifie pas la significativité des variables d'intérêt. De plus, la valeur des coefficients pour les variables significatives demeure à l'intérieur de leur intervalle de confiance³². Ainsi, ce test confirme que les résultats sont robustes par rapport à l'effet de groupe puisqu'ils ne sont pas générés par un ou quelques *wards* trop influents.

7 Discussion et conclusion

L'objectif de cette étude était d'estimer l'effet des restrictions environnementales et du développement écotouristique dans les zones protégées, sur le bien-être locales. Nous avons développé un modèle hiérarchique linéaire à deux niveaux et utilisé des données népalaises. Les résultats présentés dans cette étude appuient l'hypothèse suggérée par plusieurs auteurs : le développement écotouristique contribue au bien-être des ménages résidant à l'intérieur d'une zone protégée. En regroupant les ménages selon leur catégorie d'emploi, soit salariés ou travailleurs autonomes, nous avons montré que l'effet sur le bien-être varie entre les deux groupes. En effet, nous avons estimé que l'implication d'un ménage dans l'écotourisme comme salarié n'a pas d'effet significatif sur son bien-être. Par ailleurs, l'implication d'un ménage dans l'écotourisme comme travailleur autonome contribue significativement à l'augmentation de ses dépenses de consommation. Au Népal, les deux catégories d'emplois sont caractérisées par des conditions de travail qui diffèrent considérablement, expliquant sans doute cet écart dans les résultats. En effet, parmi les ménages interrogés, le revenu moyen par individu provenant de l'écotourisme est de 123 053 NPR pour les salariés et de 708 346 NPR pour les travailleurs autonomes. Les principaux emplois occupés par

32. Pour *tour^{aut}*, l'intervalle de confiance est de [0,30 ; 0,70] et pour *tourward^{aut}*, l'intervalle de confiance est de [0,27 ; 1,78].

les répondants salariés consistent à travailler comme guide touristique ou porteur en montagne, ou à s'acquitter de différentes tâches dans le milieu de l'hôtellerie ou de la restauration. Ces travailleurs sont généralement rémunérés sur une base quotidienne et non-régulière. Aussi, améliorer les conditions de travail des salariés semble nécessaire pour que cette catégorie d'emploi en vienne à contribuer positivement au bien-être. D'un autre côté, la majorité des ménages impliqués dans l'écotourisme comme travailleurs autonomes possède un hôtel ou un restaurant. Ces entreprises génèrent de l'emploi intra-ménage et sont plus lucratives. Puisque le gouvernement du Népal intègre le développement de l'écotourisme dans sa stratégie de réduction de la pauvreté, encourager le développement d'entreprises locales devrait être envisagé. Par exemple, faciliter l'accès au crédit pour les petites entreprises, en particulier en région éloignée où les services financiers sont parfois totalement absents (UNDP, 2015a), augmenterait la capacité d'investissement des ménages. De plus, proposer des programmes de formation courts et adaptés liés à l'entrepreneuriat, à la gestion de petites entreprises et aux services d'accueil permettrait d'améliorer les compétences requises pour développer l'offre de services écotouristiques. Enfin, le développement d'infrastructures, notamment pour améliorer le réseau routier, l'accès à l'électricité et les télécommunication permettrait d'augmenter la qualité de l'expérience écotouristique proposée afin d'en augmenter la demande.

Nos résultats montrent également que l'effet sur le bien-être d'être impliqué dans l'écotourisme comme travailleur autonome varie entre les *wards*. Nous avons calculé que cet effet est positif, pour une majorité sinon tous les *wards*. Notre résultat est cohérent avec Yergeau, Boccanfuso et Goyette (2014), qui montrent de manière théorique que l'effet de l'écotourisme sur le bien-être dans les zones protégées varie selon le contexte géographique. Les auteurs expliquent ce résultat en s'appuyant sur différentes études empiriques suggérant que l'effet de la conservation sur le bien-être dans les zones protégées dépend de caractéristiques géographiques telles que l'altitude, la dénivellation, le climat, la condition des bassins versants et la couverture forestière (Sims, 2010; Ferraro and Hanauer, 2011; Canavire-Bacarreza and Hanauer, 2013). Cependant, il n'existe pas d'information suffisamment désagrégée et disponible spécifique aux *wards*, qui nous permettrait d'expliquer la variance du coefficient associé à l'implication dans l'écotourisme comme travailleur autonome. Aussi, déterminer les facteurs causant la variation de l'effet du développement écotouristique sur le bien-être constitue un défi empirique qui par ailleurs mériterait d'être considéré dans de futures recherches. En effet, identifier les conditions qui optimisent l'effet du développement écotouristique

sur le bien-être permettrait d'améliorer la portée et l'efficacité des politiques de développement et de réduction de la pauvreté qui intègrent ce secteur d'activité.

En plus de l'effet lié à l'implication individuelle des ménages, nous avons estimé l'effet de la part des ménages appartenant à un *ward*, impliqués dans l'écotourisme, sur le bien-être moyen dans ce *ward*. Les résultats obtenus suggèrent que l'écotourisme génère une externalité positive sur le bien-être des ménages, qu'ils soient impliqués ou non dans le secteur. En effet, nous avons estimé qu'une augmentation de la part des ménages d'un *ward* impliqués dans l'écotourisme comme travailleurs autonomes de 10 points de pourcentage, produit un effet multiplicatif de 1,0549 sur la consommation des ménages appartenant à ce *ward*, indépendamment de leur implication individuelle dans le secteur. Ce résultat est important puisqu'il indique qu'une certaine part du revenu provenant de l'écotourisme est dépensée ou redistribuée localement. L'industrie du tourisme dans les pays en développement est souvent critiquée sous prétexte que les retombées économiques ne profitent qu'à quelques individus et qu'une part importante des bénéfices est redirigée hors des communautés locales. Bien que nous ne possédions pas suffisamment d'information pour déterminer la part du revenu qui demeure à l'échelle locale, nos résultats montrent tout de même que l'externalité est suffisante pour générer une hausse significative du bien-être local, même pour les ménages qui ne sont pas directement impliqués dans le secteur.

Enfin, nous montrons qu'être contraint au niveau de l'utilisation des ressources naturelles par la réglementation en vigueur dans les zones protégées n'affecte pas significativement le bien-être des ménages. Nous avons suggéré que les ménages résidant à l'intérieur des zones protégées enquêtés ont pu développer des mécanismes leur permettant de maintenir un certain niveau de consommation malgré les restrictions environnementales. Cependant, nous observons que la moyenne des dépenses de consommation des ménages qui ne sont pas contraints est de 170 728 RNP, ce qui est significativement supérieur à la consommation moyenne des ménages contraints de 140 885 RNP pour $cont = 1$ et de 138 094 RNP pour $cont = 2$ ³³. De plus, environ 70 % des ménages contraints (avec $cont = 1$ ou $cont = 2$) ont collecté du bois dans la forêt au cours des 12 derniers mois et un peu plus de 70 % de ces ménages travaillent dans le secteur agricole. Ainsi, bien que nos résultats ne montrent pas de lien direct entre la variable de restriction environnementale et le bien-être, il semble que les ménages contraints soient tout de même plus dépendants de l'utilisation des ressources naturelles et caractérisés par un niveau de bien-être inférieur. Aussi, nous interpré-

33. Significativité de l'écart à 10 %.

tons notre résultat avec précaution et réitérons l'importance d'implanter des mécanismes visant à compenser les populations locales pour les restrictions imposées sur l'utilisation des ressources naturelles, lors de la mise en place de politiques environnementales.

D'un point de vue politique, cette étude appuie la recommandation formulée par les Nations Unis, d'intégrer le secteur touristique dans les stratégies de développement et de réduction de la pauvreté. Par ailleurs, identifier les facteurs qui causent une variation de l'effet de l'écotourisme sur le bien-être, permettrait de mieux comprendre les conditions nécessaires pour optimiser les résultats visés par le développement écotouristique. De plus, les effets redistributifs générés par ce secteur ainsi que les impacts sur la pauvreté et l'inégalité devraient être étudiés. Enfin, examiner l'impact d'autres mécanismes que l'écotourisme, susceptibles de contribuer à l'augmentation du bien-être dans les zones protégées serait complémentaire à cette étude et permettrait d'approfondir la compréhension de la relation entre conservation environnementale et bien-être.

Références

- AASSVE, A. et ARPINO, B. (2007). Dynamic multi-level analysis of households' living standards and poverty : Evidence from Vietnam. ISER Working Paper 2007-10.
- ACHARYA, H. (2014). Overview of protected areas management in Nepal. Document consulté en juin 2016 de <https://www.cbd.int/doc/meetings/ecr/cbwecr-2014-07/other/cbwecr-2014-07-presentation-day2-04-en.pdf>.
- ADAMS, W., AVELING, R., BROCKINGTON, D., DICKSON, B., ELLIOTT, J., HUTTON, J., ROE, D., VIRA, B. et WOLMER, W. (2004). Biodiversity conservation and the eradication of poverty. *Science*, 36:1146–1148.
- ADHIKARI, B. (2005). Poverty, property rights and collective action : Understanding the distributive aspects of common property resources management. *Environment and Development Economics*, 10:7–31.
- AMATYA, S. (2013). Financing for sustainable forest management in Nepal. Indufor.
- ANDAM, K., FERRARO, P., SIMS, K., HEALY, A. et HOLLAND, W. (2010). Protected areas reduced poverty in Costa Rica and Thailand. *Proceedings of the National Academy of Sciences of the United States of America*, 107(22):9996–10001.
- ANGELSON, A., JAGGER, P., BABIGUMIRA, R., BELCHER, B., HOGARTH, N., BAUCH, S., BÖRNER, J., SMITH-HALL, C. et WUNDER, S. (2014). Environmental income and rural livelihoods : A global-comparative analysis. *World Development*, 64:S12–S28.
- ARCHER, K., LEMESHOW, S. et HOSMER, D. (2007). Goodness-of-fit for logistic regression models when data are collected using a complex sampling design. *Computational Statistics and Data Analysis*, 51:4450–4464.
- ASIAN DEVELOPMENT BANK (ADB). (2016). Poverty in Nepal. Page consulté en juin 2016 de <http://www.adb.org/countries/nepal/poverty>.
- ASPAROUHOV, T. (2006). General multi-level modeling with sampling weights. *Communication in statistics - Theory and methods*, 35:439–460.
- ASPAROUHOV, T. et MUTHEN, B. (2005). Multivariate statistical modeling with survey data. Proceedings for Federal Committee on Statistical Methodology research conference.
- BALAND, J., BARDHAN, P., DAS, S., MOOKHERJEE, D. et SARKAR, R. (2010). The environmental impact of poverty : Evidence from firewood collection in rural Nepal. *Economic Development and Cultural Change*, 59(1):23–61.
- BANQUE MONDIALE (BM). (2011). Defining welfare measures. Page consultée en mars 2013 de <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTPA/0,,contentMDK:20242876~menuPK:435055~pagePK:148956~piPK:216618~theSitePK:430367~isCURL:Y~isCURL:Y,00.html>.
- BARAL, N. (2013). Evaluation and resilience of ecotourism in the Annapurna Conservation Area, Nepal. *Environmental Conservation*, 41(1):84–92.

- BARAL, N., STERN, M. J. et HEINEN, J. T. (2007). Integrated conservation and development project life cycles in the Annapurna Conservation Area, Nepal : Is development overpowering conservation? *Biodiversity Conservation*.
- BASNETT, Y., HENLEY, G., HOWELL, J., JONES, H., LEMMA, A. et PANDEY, P. (2014). Structural economic transformation in Nepal. A diagnostic study submitted to DFID Nepal. Document consulté en ligne de <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9019.pdf>.
- BELSLEY, D., KUH, E. et WELSCH, R. (1980). *Regression diagnostics : Identifying influential data and sources of collinearity*. John Wiley, New York.
- BHATTA, S. et SHARMA, S. (2006). The determinants and consequences of chronic and transient poverty in Nepal. CPRC Working paper 66.
- BIRDLIFE INTERNATIONAL (BLI) (2016). Threats to biodiversity have much deeper causes. Page consultée en juin 2016 de <http://www.birdlife.org/datazone/sowb/pressure/PRESS0>.
- BROCKINGTON, D. et WILKIE, D. (2015). Protected areas and poverty. *Phil. Trans. R. Soc. B*, 370(20140271):1–6. <http://dx.doi.org/10.1098/rstb.2014.0271>.
- BRUMBACK, B., ZHENG, H. et DAILEY, A. (2013). Adjusting for confounding by neighborhood using generalized linear mixed models and complex survey data. *Statistics in Medicine*, 32:1313–1324.
- BUDHATHOKI, P. (2005). Landscape conservation initiatives in Nepal : Opportunities and challenge. Page consulté en juin 2016 de <https://portals.iucn.org/library/efiles/html/PA-protected-landscape-approach/PartII-section6.html>.
- CAMERON, C. et MILLER, D. (2010). Robust inference with clustered data. Working paper, University of California, Department of Economics, no. 10.7.
- CAMERON, C. et TRIVEDI, P. (2005). *Microeconometrics. Methods and applications*. Cambridge University Press, New York.
- CANAVIRE-BACARREZA, G. et HANAUER, M. (2013). Estimating the impacts of Bolivia’s protected areas on poverty. *World Development*, 41:265–285.
- CARLE, A. (2009). Fitting multilevel models in complex survey data with design weights : Recommendations. *BMC Medical research methodology*, 9(1):1.
- CENTRAL BUREAU OF STATISTICS (CBS). (2011). Nepal living standard survey. Statistical Report.
- CENTRAL BUREAU OF STATISTICS (CBS). (2012). National population census 2011.
- CHEN, K. et WANG, T. (2015). Determinants of poverty status in Taiwan : A multilevel approach. *Social Indicators Research*, 123:371–389.
- CHOK, S., MACBETH, J. et WARREN, C. (2007). Tourism as a tool for poverty alleviation : A critical analysis of pro-poor tourism and implications for sustainability. *Current Issues in Tourism*, 10(2):144–165.

- COAD, L., CAMPBELL, A., MILES, L. et HUMPHRIES, K. (2008). The costs and benefits of forest protected areas for local livelihoods : A review of the current literature. Working Paper. UNEP World Conservation Monitoring Centre, Cambridge, U.K.
- DEATON, A. (1997). *The analysis of household surveys : A microeconomic approach to development policy*. The World Bank.
- DEATON, A. et ZAIDI, S. (2003). *Guidelines for construction consumption aggregates for welfare analysis*. The World Bank, Washington D.C. LSMS working paper. Number 135.
- DUDLEY, N., MANSOURIAN, S., STOLTON, S. et SUKSUWAN, S. (2008). Safety Net. Protected areas and poverty reduction. Rapport technique, World Wide Fund for Nature and Equilibrium. ISBN : 978-2-88085-279-5.
- EUGENIO, M., MORALES, N. et SCARPA, R. (2004). Tourism and economic growth in Latin American countries : A panel data approach. *Natural Resources Management*, pages 1–20.
- FERRARO, P. et HANAUER, M. (2011). Protecting ecosystems and alleviating poverty with parks and reserves : Win-win or tradeoffs ? *Environmental and resource economics*, 48:269–286.
- FERRARO, P. et HANAUER, M. (2014). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Science of the United States*, 111:4332–4337.
- FERRARO, P., HANAUER, M. et SIMS, K. (2011). Conditions associated with protected area success in conservation and poverty reduction. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34):13913–13918.
- GOLDSTEIN, H. et THOMAS, S. (1996). Using examinations results as indicators of school and college performance. *Journal of the Royal Statistical Society. Series A*, 159:149–163.
- GOODWIN, H. (2006). Measuring and reporting the impact of tourism on poverty. *In Cutting edge research in tourism – New directions, challenges and applications*. School of Management, University of Surrey, UK, 6 au 9 juin.
- GOVERNMENT OF NEPAL (DNPWC) (2014). Protected areas. <http://www.dnpwc.gov.np/>.
- GRIEG-GRAN, M., PORRAS, I. et WUNDER, S. (2005). How can market mechanisms for forest environmental services help the poor ? Preliminary lessons from latin america. *World Development*, 33(9):1511–1527.
- GRILLI, L. et PRATESI, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs. *Survey methodology*, 30(1):93–104.
- HEINEN, J. et SHRESTHA, S. (2006). Evolving policies for conservation : An historical profile of the protected area system of Nepal. *Journal of Environmental Planning and Management*, 49(1):41–58.
- HOLDEN, A. et SPARROWHAWK, J. (2002). Understanding the motivation of ecotourists : the case of trekkers in Annapurna, Nepal. *International Journal of Tourism Research*, 4(6):435–446.

- HOX, J. (1998). *Classification, data analysis, and data highways*, chapitre Multilevel modeling : When and why ?, pages 147–154. Springer Verlag.
- HOX, J. (2010). *Multilevel analysis. Techniques and applications*. Routledge, Great-Britain, second édition.
- INTERNATIONAL FUND FOR AGRICULTURAL DEVELOPMENT (IFAD). (2013). Enabling poor rural people to overcome poverty in Nepal. Document consulté en ligne en juin 2016 de <https://www.ifad.org/documents/10180/c3f05a7f-627b-40a6-8bce-c3330c9849dc>.
- INTERNATIONAL FUND FOR AGRICULTURAL DEVELOPMENT (IFAD). (2014). Rural poverty in Nepal. Page consultée en février 2014 de <http://www.ruralpovertyportal.org/country/home/tags/nepal>.
- INTERNATIONAL LABOUR ORGANIZATION (ILO). (2005). Dalits and labour in Nepal : Discrimination and forced labour. Document consulté en avril 2016 de http://www.ilo.org/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-kathmandu/documents/publication/wcms_112922.pdf.
- INTERNATIONAL UNION FOR CONSERVATION OF NATURE (IUCN). (2016). Protected Areas Categories. Page consulté en juillet 2016 de <https://www.iucn.org/theme/protected-areas/about/protected-areas-categories>.
- JIA, Y., STOKES, L., HARRIS, I. et WANG, Y. (2011). Performance of random effects model estimators under complex sampling designs. *Journal of educational and behavioral statistics*, 36(1):6–32.
- KARANTH, K. et NEPAL, S. (2011). Local residents perception of benefits and losses from protected areas in India and Nepal. *Environmental Management*.
- KEITER, R. (1995). Preserving Nepal’s National Parks : Law and Conservation in the Developing World. *Ecology Law Quarterly*, 22(3):591–675.
- KIM, K., LEE, Y. et LEE, Y. (2010). A multilevel analysis of factors related to poverty in welfare states. *Social Indicators Research*, 99(3):391–404.
- KIM, R., MOHANTY, S. et SUBRAMANIAN, S. (2016). Multilevel geographies of poverty in India. *World Development*, 87:349–359.
- KORN, E. et GRAUBARD, B. (2003). Estimating variance components by using survey data. *Journal of the Royal Statistical Society. Series B*, 65(1):175–190.
- KOVACEVIC, M. et RAI, S. (2003). A pseudo maximum likelihood approach to multilevel modelling of survey data. *Communications in statistics. Theory and methods*, 32(1):103–121.
- LAIRD, N. et WARE, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.
- LANLY, J.-P. (2003). Deforestation and forest degradation factors. Page consultée en juillet 2016 de <http://www.fao.org/docrep/article/wfc/xii/ms12a-e.htm>.

- LOKSHIN, M., BONTCH-OSMOLOVSKI, M. et GLINSKAYA, E. (2010). Work-related migration and poverty reduction in Nepal. *Review of Development Economics*, 14(2):323–332.
- MAAS, C. et HOX, J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3):86–92.
- MEHTA, J. et KELLERT, S. (1998). Local attitudes toward community-based conservation policy and programmes in Nepal : a case study in the Makalu-Barun Conservation Area. *Environmental Conservation*, 25(4):320–333.
- MENG, F., LI, X. et UYSAL, M. (2010). Tourism development and regional quality of life : The case of China. *Journal of China Tourism Research*, 6(2):164–182.
- MÖHRING, K. et SCHMIDT, A. (2012). Multi level tools. Influential cases in multi level modeling. Document consulté en juillet 2016 de http://www.stata.com/meeting/germany12/abstracts/desug12_moehring.pdf.
- MINISTRY OF CULTURE, TOURISM AND AVIATION (MTCA). (2013). Nepal tourism statistics 2012.
- MINISTRY OF CULTURE, TOURISM AND AVIATION (MTCA). (2015). Nepal tourism statistics 2014.
- NAGENDRA, H. et GOKHALE, Y. (2008). Management regimes, property rights and forest biodiversity in Nepal and India. *Environmental Management*, 49:719–733.
- NEPAL, K. (2000). Tourism in protected areas. The Nepalese Himalaya. *Annals of Tourism Research*, 27(3):661–681.
- NEPAL, K. (2002). Mountain ecotourism and sustainable development. Ecology, economics and ethics. *Mountain Research and Development*, 22(2):104–109.
- NETO, F. (2003). A new approach to sustainable tourism development : Moving beyond environmental protection. *Natural Resources Forum*, 27:212–222.
- NOWAK, J., SAHLI, M. et SGRO, P. (2003). Tourism, trade and domestic welfare. *Pacific economic review*, 8:245–258.
- NYAUPANE, G. et THAPA, B. (2004). Evaluation of Ecotourism : A Comparative Assessment in the Annapurna Conservation Area Project, Nepal. *Journal of Ecotourism*, 3(1):20–45.
- PACCAGNELLA, O. (2006). Centering or not centering in multilevel models ? The role of group mean and the assessment of group effects. *Evaluation review*, 30(1):66–85.
- PFEFFERMANN, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61:317–337.
- PFEFFERMANN, D., SKINNER, C., HOLMES, D., GOLDSTEIN, H. et RASBASH, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society. Series B*, 60(1):23–40.

- POTTHOFF, R., WOODBURY, M. et MANTON, K. (1992). "Equivalent sample size" and "Equivalent degree of freedom" refinements for inference using survey weights under superpopulation models. *Journal of the American Statistical Association*, 87(418):383–396.
- RABE-HESKETH, S. et SKRONDAL, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society. Series A*, 169(4):805–827.
- RICHARDSON, R., FERNANDEZ, A., TSCHIRLEY, D. et TEMBO, G. (2012). Wildlife conservation in Zambia : Impacts on rural household welfare. *World Development*, 40(5):1068–1081.
- ROBALINO, J. (2007). Land conservation policies and income distribution : Who bears the burden of our environmental efforts ? *Environment and development economics*, 12:521–533.
- ROBALINO, J. et VILLALOBOS-FIATT, L. (2015). Protected areas and economic welfare : An impact evaluation of national parks on local workers' wages in Costa Rica. *Environment and Development Economics*, 20(3):283–310.
- ROBINSON, E., ALBERS, H. et WILLIAMS, J. (2008). Spatial and temporal aspects of non-timber forest product extraction : the role of community resource management. *Journal of environmental economics and management*, 56:234–245.
- ROBINSON, E. et LOKINA, R. (2011). A spatial-temporal analysis of the impact of access restrictions on forest landscapes and household welfare in tanzania. *Forest policy and economics*, 13:79–85.
- RURAL RECONSTRUCTION NEPAL (RRN) et CENTRE FOR INTERNATIONAL STUDIES AND CO-OPERATION (CECI-NEPAL). (2007). Development strategies for a new Nepal : A national debate. Document consulté en juillet 2007 de http://www.rrn.org.np/wp-content/uploads/2013/07/new_nepal_report.pdf.
- SEQUEIRA, N. et NUNES, P. (2008). Does tourism influence economic growth ? A dynamic panel data approach. *Applied Economics*, 40:2431–2441.
- SHRESTHA, H. (2016). Visitor numbers in 2015 not so low but tourism suffered. Page consultée en juillet 2016 de <http://www.nepalmountainnews.com/cms/2016/01/12/visitor-numbers-in-2015-not-so-low-but-tourism-suffered/>.
- SHRESTHA, U., SHRESTHA, S., CHAUDHARY, P. et CHAUDHARY, R. (2010). How representative is the protected areas system of Nepal ? *Mountain Research and Development*, 30(3):282–294.
- SIMS, K. (2010). Conservation and development : Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, 60(2):94–114.
- SKINNER, C. (1989). *Analysis of complex surveys*, chapitre Domain means, regression and multivariate analysis, pages 59–87. Wiley, Chichester.
- SNIJDERS, T. et BERKHOF, J. (2008). *Handbook of multilevel analysis*, chapitre Diagnostic checks for multilevel models, pages 141–175. Springer, New York.
- SNIJDERS, T. et BOSKER, R. (2012). *Multilevel Analysis. An introduction to basic and advanced multilevel modeling*. Sage Publication, second édition.

- SPITERI, A. et NEPAL, S. (2008). Distributing conservation incentives in the buffer zone of Chitwan National Park, Nepal. *Environmental Conservation*, 35(1):76–86.
- STEELE, F., VIGNOLES, A. et JENKINS, A. (2007). The effect of school resources on pupil attainment : a multilevel simultaneous equation modelling approach. *Journal of the Royal Statistical Society. Series A*, 170(3):801–824.
- STEENBERGEN, M. et JONES, B. (2002). Modeling multilevel data structures. *American Journal of Political Science*, pages 218–237.
- UNITED NATIONS DEVELOPMENT PROGRAM (UNDP). (2015a). Enhancing access to financial services. Document consulté en novembre 2016 de <http://www.np.undp.org/content/nepal/en/home/operations/projects/closed-projects/poverty-reduction/eafs.html>.
- UNITED NATIONS DEVELOPMENT PROGRAM (UNDP). (2015b). Rapport sur le développementn humain 2015. Document consulté en juillet 2016 de http://hdr.undp.org/sites/default/files/2015_human_development_report_overview_-_fr.pdf.
- UNITED NATIONS ENVIRONMENT PROGRAM (UNEP) (2014). Mapping the world’s special places. Page consultée en juillet 2016 de <http://www.unep-wcmc.org/featured-projects/mapping-the-worlds-special-places>.
- UNITED NATIONS (UN) (2016). Objectifs du développement durable. Page consultée en juin 2016 de <http://www.un.org/sustainabledevelopment/fr/biodiversity/>.
- WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica : Journal of the Econometric Society*, pages 817–838.
- WOODLRIDGE, J. (2003). Cluster-sample methods in applied econometrics. *The American Economic Review*, 93(2):133–138.
- WORLD BANK. (2016). The Nepal earthquakes of 2015 : One year on. Page consultée en juillet 2016 de <http://blogs.worldbank.org/endpovertyinsouthasia/nepal-earthquakes-2015-one-year>.
- WORLD TOURISM ORGANIZATION (WTO). (2010). Le marché italien de l’écotourisme. Page consultée en novembre 2012 de <http://pub.world-tourism.org/WebRoot/Store/Shops/Infoshop/Products/1251/1251-2.pdf>.
- WORLD TOURISM ORGANIZATION (WTO). (2013). UN General Assembly : Eco-tourism key to eradicating poverty and protecting environment. Page consultée en juillet 2016 de <http://media.unwto.org/press-release/2013-01-03/un-general-assembly-ecotourism-key-eradicating-poverty-and-protecting-envir>.
- WORLD TRAVEL AND TOURISM COUNCIL (WTTC). (2015). Travel and tourism. Economic impact 2015. Nepal.
- YERGEAU, M. (2015). Conservation, écotourisme et bien-être : leçons népalaises. *Revue d’économie du développement*, 23(1):29–165.
- YERGEAU, M., BOCCANFUSO, D. et GOYETTE, J. (2014). Conservation and welfare : Toward a reconciliation of theory and facts. GRÉDI Working paper 14-05.

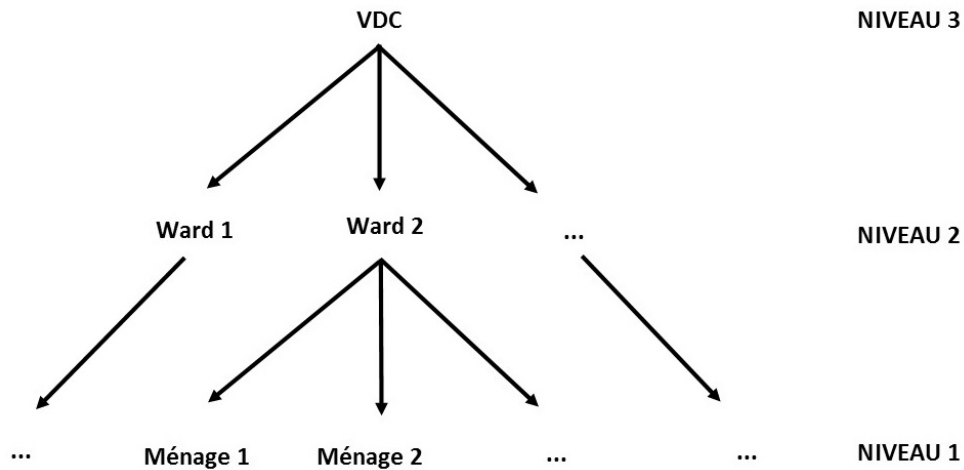
YUNIS, E. (2004). Chief sustainable development of tourism, World Tourism Organization. Document consulté en mars 2011 de http://www.regione.toscana.it/regione/multimedia/RT/documents/1219927391372_eng_yunis.pdf.

A Résumé de la structure de la base de données

TABLE 7 – Résumé de la structure de la base de données

	VDC sélectionnés	Nombre de <i>wards</i> inclus dans l'échantillon	Nombre de ménages sélectionnés
Annapurna	Lete	8	149
	Narchyang	6	117
	Lumle	6	120
	Ghandruk	7	150
Langtang	Dhunche	8	174
	Syafu	7	161
	Laharepauwa	5	156
Chitwan	Meghauly	9	177
	Dibyapuri	6	179
	Bachhauri	9	180
Niveau d'analyse	Niveau 3	Niveau 2	Niveau 1

Figure 2. Structure hiérarchique de la base de données



B Méthode d'estimation

Soit le modèle :

$$y_{ij} = x_{ij}\beta + z_{ij}u_j + \epsilon_{ij} \quad (17)$$

$$\text{Var} \begin{pmatrix} U \\ \epsilon \end{pmatrix} = \begin{pmatrix} \Sigma & 0 \\ 0 & \sigma^2 \end{pmatrix} \quad (18)$$

tel que définit dans (1) et (2), avec i l'unité du niveau 1 et j l'unité du niveau 2.

Soit $f(y_{ij}|u_j, \beta, \sigma^2)$ la densité marginale de y , conditionnelle aux paramètres estimés β , σ^2 et à l'effet aléatoire u_j . Puisque les variables y_{ij} et u_j ne sont pas indépendantes, utilisons la densité jointe :

$$f(y_{ij}, u_j|\beta, \sigma^2, \Sigma) = f(y_{ij}|u_j, \beta, \sigma^2)f(u_j|\Sigma) \quad (19)$$

et intégrons pour obtenir la densité marginale de y , inconditionnelle à l'effet aléatoire :

$$f(y_{ij}|\beta, \sigma^2, \Sigma) = \int f(y_{ij}|u_j, \beta, \sigma^2)f(u_j|\Sigma)du_j \quad (20)$$

Soient J le nombre d'unités de niveau 2 (cluster) et n_j le nombre d'unités de niveau 1 par cluster. Puisque $y_{ij} \perp y_{i'j}$, $\forall i, i' \in j$ et que $u_j \perp u_{j'}$, $\forall j, j'$, alors la densité jointe des observations sur l'ensemble de la distribution s'écrit :

$$f(y|\beta, \sigma^2, \Sigma) = \prod_{j=1}^J \int \prod_{i=1}^{n_j} f(y_{ij}|u_j, \beta, \sigma^2)f(u_j|\Sigma)du_j \quad (21)$$

On obtient ainsi la fonction de vraisemblance $\mathcal{L}(\beta, \sigma^2, \Sigma|y) = f(y|\beta, \sigma^2, \Sigma)$. Prenons le logarithme :

$$l(\beta, \sigma^2, \Sigma|y) = \log \left[\prod_{j=1}^J \int \prod_{i=1}^{n_j} f(y_{ij}|u_j, \beta, \sigma^2)f(u_j|\Sigma)du_j \right] \quad (22)$$

Redéfinissons :

$$\prod_{i=1}^{n_j} f(y_{ij}|u_j, \beta, \sigma^2) = \exp \left[\sum_{i=1}^{n_j} \log f(y_{ij}|u_j, \beta, \sigma^2) \right] \quad (23)$$

En insérant (23) dans (22) :

$$l(\beta, \sigma^2, \Sigma|y) = \sum_{j=1}^J \log \left(\int \exp \left[\sum_{i=1}^{n_j} \log f(y_{ij}|u_j, \beta, \sigma^2) \right] f(u_j|\Sigma)du_j \right) \quad (24)$$

Soient ω_j , le poids attribué au cluster j , correspondant à l'inverse de sa probabilité de sélection et ω_{ij} , le poids attribué à l'unité i dans le cluster j , correspondant à l'inverse de la probabilité de sélection de l'unité i , conditionnelle à la sélection du cluster j . En intégrant dans (24) les poids d'échantillonnage tel que dans Rabe-Hesketh et Skrondal (2006), nous obtenons la fonction de log-vraisemblance suivante :

$$l(\beta, \sigma^2, \Sigma|y) = \sum_{j=1}^J \omega_j \log \left(\int \exp \left[\sum_{i=1}^{n_j} \omega_{ij} \log f(y_{ij}|u_j, \beta, \sigma^2) \right] f(u_j|\Sigma) du_j \right) \quad (25)$$

La fonction de log-vraisemblance estimée $l(\beta, \sigma^2, \Sigma|y)$ est donc composée de la somme pondérée de J fonctions de log-vraisemblance indépendantes.

C Paramétrisation

Afin de reproduire le plan d'échantillonnage ainsi que la taille des clusters, nous répliquons les observations contenues dans la base de données en utilisant les poids d'échantillonnage de telle sorte que le nombre d'observations soit égal à la taille de la population réelle. Puis, nous procédons à un tirage aléatoire selon les probabilités de sélection utilisées lors de la collecte de données afin que l'échantillon tiré soit de taille égale à l'échantillon réel. Pour reproduire la corrélation intra-classe, nous simulons un vecteur d'effets aléatoires u et de résidus e de telle sorte que $u \sim \mathcal{N}(0, 0.1)$ et $e \sim \mathcal{N}(0, 0.4)$. Cette paramétrisation tend vers celle associée à l'estimation du modèle (11) et fait en sorte que la corrélation intra-classe $ICC = 0.2$, ce qui tend aussi vers celle calculée à partir des résultats de (11). Deux variables explicatives sont intégrées au modèle simulé. La première x_{ij} est une variable du niveau ménage, alors que la seconde x_j est du niveau *ward*. Nous motivons le choix d'utiliser ces deux variables par le fait que la performance des estimateurs peut différer entre les coefficients des différents niveaux (Rabe-Hesketh et Skrondal, 2006). Mentionnons que ces variables ne sont pas simulées mais construites à partir de la base de données originale. La variable dépendante y est ensuite calculée tel que $y_{ij} = 1 + x_{ij} + x_j + u_j + e_{ij}$, ce qui impose des coefficients $\beta_0 = \beta_{x_{ij}} = \beta_{x_j} = 1$. Enfin, l'informativité des poids par rapport à la variable dépendante est un paramètre difficile à reproduire, puisque difficile à mesurer³⁴. Aussi, nous nous appuyons sur les similitudes entre les paramètres simulés et estimés, de même que sur l'utilisation des poids d'échantillonnage et de variables explicatives issus de la base originale, pour argumenter que l'informativité des poids dans l'échantillon simulé sera semblable à celle de la base de données originale.

34. Voir Pfeffermann (1993) pour plus de détails sur l'informativité des poids.

D Test de la distance de Cook

TABLE 8 – Résultats du test de la distance de Cook

Ward ID	C_j^F	C_j^A	C_j
1	2,825	0,046	2,612
2	2,003	0,010	1,850
3	0,916	0,041	0,849
4	0,577	0,025	0,535
5	0,549	0,023	0,508
6	0,504	0,033	0,468
7	0,456	0,014	0,422
8	0,346	0,004	0,320
9	0,285	0,022	0,265
10	0,226	0,709	0,263
11	0,271	0,001	0,250
12	0,244	0,019	0,226
13	0,209	0,024	0,195
14	0,201	0,119	0,194
15	0,207	0,039	0,194
16	0,190	0,063	0,180
17	0,188	0,071	0,179
18	0,160	0,010	0,148
19	0,157	0,043	0,148
20	0,114	0,016	0,107
21	0,102	0,091	0,101
22	0,092	0,042	0,088
23	0,083	0,126	0,086
24	0,086	0,013	0,081
25	0,077	0,017	0,073
26	0,078	0,007	0,073
27	0,071	0,006	0,066
28	0,060	0,041	0,058
29	0,057	0,049	0,057
30	0,024	0,061	0,027
31	0,012	0,088	0,018

Seuil critique = 0,06