

**Zahlungsbereitschaftsanalysen in der  
umweltökonomischen Bewertung von öffentlichen Gütern des Waldes**

Dissertation  
zur Erlangung des Doktorgrades  
der Fakultät für Forstwissenschaften und Waldökologie  
der Georg-August-Universität Göttingen

vorgelegt von  
Priska Johanna Weller  
geboren in Frankfurt am Main

Göttingen, 2018

1. Gutachter: Prof. Dr. Matthias Dieter
2. Gutachter: Prof. Dr. Bernhard Möhring

Tag der mündlichen Prüfung: 18. Oktober 2018

## **Inhaltsverzeichnis**

Zusammenfassung .....	i
Abbildungs- und Tabellenverzeichnis .....	v
1. Einführung in die wissenschaftliche Fragestellung .....	1
Einleitung .....	1
Theoretischer Hintergrund und Methode .....	2
Überblick über Literatur und forstpolitische Programme .....	10
Forschungsfragen .....	13
2. Einordnung der eigenen Ergebnisse in den fachlichen Kontext .....	15
2.1 Erster Artikel „Aktuelle und potentielle Erholungsleistung der Wälder in Deutschland: Monetärer Nutzen der Erholung im Wald aus Sicht der Bevölkerung“ (Elsasser und Weller, 2013, AFJZ) .....	16
Zusammenfassung .....	16
Diskussion und Einordnung in den thematischen Rahmen .....	18
2.2 Zweiter Artikel „Preferences for forest structural attributes in Germany – evidence from a Choice Experiment“ (Weller und Elsasser, 2018, Forest Policy and Economics) .....	19
Zusammenfassung .....	19
Diskussion und Einordnung in den thematischen Rahmen .....	21
2.3 Dritter Artikel „Stated and inferred attribute non-attendance in a design of designs approach“ (Weller, Oehlmann, Mariel, Meyerhoff, 2014, JoCM) .....	22
Zusammenfassung .....	23
Diskussion und Einordnung in den thematischen Rahmen .....	27
2.4 Vierter Artikel „The Influence of Design Dimensions on Stated Choices in an Environmental Context“ (Meyerhoff, Oehlmann, Weller, 2015, ERE) .....	29
Zusammenfassung .....	29
Diskussion und Einordnung in den thematischen Rahmen .....	31
3. Diskussion und Schlussfolgerungen .....	33
Literaturverzeichnis .....	38
Anhang: wissenschaftliche Artikel .....	48
Danksagung .....	vi
Eidesstattliche Erklärung .....	vii

## **Zusammenfassung**

Durch die mangelnde Verfügbarkeit von Marktpreisen für öffentliche Güter des Waldes entsteht der Bedarf, verlässliche Bewertungen für diese zu erlangen. Das Marktversagen, welches entsteht wenn Güter Dritte positiv oder negativ beeinflussen, aber nicht auf Märkten gehandelt werden, kann nur so überwunden werden. Die Bewertung nicht-marktlicher Güter erfolgt über Methoden der geäußerten Präferenzen, die Nutzennäherungen aus den Aussagen von befragten Personen ableiten, indem Märkte imitiert werden. Zwei verbreitete Methoden sind dabei die bedingte Bewertung (englisch: Contingent Valuation Method, CVM) und diskrete Auswahlexperimente (englisch: Discrete Choice Experiments, DCE). Methoden der geäußerten Präferenzen versetzen die Befragten über den experimentellen Versuchsaufbau in eine Situation, in der sie ähnlich wie auf einem Markt Entscheidungen fällen sollen oder eine Zahlungsbereitschaft (ZB) äußern.

In der vorliegenden Arbeit werden drei Forschungsfragen hauptsächlich behandelt.

Zum ersten soll die kritische Anwendung der Methoden CVM und DCE die Ableitung von Bewertungsergebnissen für einzelne Leistungen des Waldes ermöglichen, zum zweiten sollen methodische Aspekte der DCE untersucht werden, die sich auf das Antwortverhalten der Befragten in Abhängigkeit von der Komplexität des Befragungsdesigns beziehen, und zum dritten soll so die Möglichkeit geschaffen werden, Güter in Verhandlungen zu positionieren und so Marktversagen teilweise zu überwinden.

Elsasser und Weller (2013) bewerten mittels CVM in einer allgemeinen Bevölkerungsbefragung das Recht, den Wald in der Wohnumgebung ein Jahr lang zur Erholung betreten zu dürfen. Die Befragten werden dafür in einer hypothetischen Marktsituation direkt nach ihrer maximalen individuellen ZB gefragt. So ergibt sich für die persönlich Befragten eine ZB i.H.v. 27€/p/a für das Waldbetretensrecht. Um die Obergrenze der ZB abschätzen zu können, wurde im Anschluss an die Frage nach der ZB nach gewünschten Verbesserungen im Wald und resultierender ZB dafür gefragt. Dies ergab, dass über die Hälfte der Befragten sich Änderungen an der Infrastruktur wünscht, ein gutes Viertel der Befragten wünscht sich Änderungen hin zu mehr Naturschutz im Wald und das verbleibende knappe Fünftel hat keine Verbesserungswünsche. Allerdings entstehen auch nach der hypothetischen Realisierung dieser Wünsche kaum zusätzliche Waldbesuche der Befragten, sowie kaum zusätzliche ZB. Die Durchführung der Befragung erfolgte in zwei separaten Stichproben, eine wurde persönlich befragt und die andere wurde in einem online-

Fragebogen befragt. Dabei fand sich kein signifikanter Unterschied in den Zahlungsbereitschaften der beiden Stichproben.

Weller und Elsasser (2018) bewerten mittels einer weiteren Bevölkerungsbefragung acht verschiedene Waldeleistungen. Das zu bewertende Gut ist eine Landschaft, charakterisiert durch unterschiedliche Merkmale, die sich auf Eigenschaften des Waldes und teilweise auch der Landschaft beziehen. Fragebogen und DCE wurde online von den Befragten beantwortet und die Befragten waren auf zwei fast gleich große Stichproben verteilt.

Die beiden verwendeten Modelle (Multinomiales Logit-Modell und Mixed Logit-Modell mit Verteilungsannahmen der ZB („WTP space“)) zeigen sehr ähnliche Präferenzen der Befragten über Änderungen an den präsentierten Waldmerkmalen. Insbesondere bei der Betrachtung der Vorzeichen, also ob positive oder negative Präferenzen vorliegen, sind die Ergebnisse fast deckungsgleich. Sichtbar wird allerdings auch, eine große Heterogenität in den Antworten der Befragten. Dies bedeutet, dass es nicht eine klar vorherrschende Meinung gibt, sondern eine Bandbreite.

Nur wenige der angebotenen Änderungen weg vom Status Quo der Landschaft generieren positive jährliche ZB. Dies gilt für eine Steigerung des Anteils von Waldflächen, jede Steigerung der Biodiversität, eine zeitliche Verschiebung der Holzernte um 20 Jahre und eine Aufhebung der Nutzung von Wald auf 10% der Flächen. Die verbleibenden Änderungen weg vom Status Quo der Landschaft sind mit unterschiedlich hohen negativen ZB belegt. Dies gilt für eine Reduktion des Anteils von Waldflächen, Vergrößerungen und Verkleinerungen des Feld-/Wald-Mosaiks, die Reduktion von Unterwuchs, auf 70% veränderte Anteile von Nadelbäumen und Erhöhungen wie Reduktionen von nichtheimischen Bäumen, sowie eine Abschaffung von ungenutzten Waldflächen. Andere Veränderungen haben nicht zu statistisch signifikanten ZB geführt: eine Veränderung des Nadelbaumanteils auf 30%, eine Vergrößerung der Unterwuchsmenge und eine größere zeitliche Verschiebung der Holzernte um 30 Jahre.

Weller, Oehlmann, Mariel und Meyerhoff (2014) untersuchen die Frage, ob die Berücksichtigung aller im DCE angebotenen Merkmale durch die Befragten von der Komplexität des Auswahl-experiments abhängen, oder ob die Befragten (vereinfachende) Antwortheuristiken verwenden, die die Entscheidung für eine Alternative erleichtern. Die Komplexität bezieht sich auf die Anzahl der Auswahlkarten, der Merkmale, der angebotenen Alternativen, der Anzahl der Merkmalsausprägungen und der Spannweite der

Merkmalsausprägungen. Zur Erhebung der geäußerten Nicht-Berücksichtigung wurden die Befragten nach dem DCE aufgefordert, auf einer fünfstufigen Skala (niemals, selten, manchmal, häufig, immer berücksichtigt) ihre persönliche Berücksichtigung für jedes der einzelnen Merkmale anzugeben. Andererseits wurden zur rechnerischen Herleitung der Nicht-Berücksichtigung von einzelnen Merkmalen fünf mögliche Klassen definiert und die Wahrscheinlichkeiten für die jeweilige Klassenzugehörigkeit mittels latentem Klassenmodell für jeden Befragten in jeder der 16 Stichproben modelliert. In den Klassen wurden 1) alle Merkmale berücksichtigt, 2) kein Merkmal berücksichtigt, 3) – 5) jeweils eins der Merkmale Waldanteil, Flächenumwandlung und Preismerkmal nicht berücksichtigt. Im Ergebnis zeigt sich, dass die Anzahl an angebotenen Alternativen und die Anzahl der Auswahlkarten Einflüsse auf die Berücksichtigung der Merkmale haben. Je höher die Anzahl der Alternativen ist, desto mehr sinkt die Berücksichtigung aller Merkmale. Und je mehr Auswahlkarten angeboten werden, desto stärker nimmt die Nicht-Berücksichtigung aller Merkmale zu.

Meyerhoff, Oehlmann und Weller (2015) untersuchen unter Verwendung des Datensatzes aus dem vorherigen Artikel (Weller et al., 2014) ob die Komplexität des Auswahl-experiments Auswirkungen auf die Abbruchrate der Befragung, sowie die Fehlervarianz der Antworten der Befragten hat. Die Komplexität bezieht sich auf die Anzahl der Auswahlkarten, der Merkmale, der angebotenen Alternativen, der Anzahl der Merkmalsausprägungen und der Spannbreite der Merkmalsausprägungen. Die Abbruchrate steigt in Abhängigkeit von bestimmten soziodemographischen Charakteristika. Außerdem steigt sie mit steigender Anzahl von Auswahlkarten, Merkmalen und Designs mit fünf Alternativen. Die angegebenen Entscheidungen werden ebenfalls von soziodemographischen Charakteristika beeinflusst, sowie allen betrachteten Komplexitätscharakteristika. Sichtbar werden hier Lerneffekte, die die Fehlervarianz der Entscheidungen reduzieren, während später im Verlauf der Beantwortung der Auswahlkarten Ermüdung einsetzt. Die aus den Modellen errechneten Koeffizienten deuten darauf hin, dass die Befragten im Durchschnitt einen größeren Waldanteil sowie höhere Werte beim Naturschutzindikator bevorzugen, während sie weniger Flächenumwandlung bevorzugen. Insgesamt zeigt sich auch, dass die ermittelten Zahlungsbereitschaftswerte sich nicht stark verändern, wenn der Einfluss der Designdimensionen auf die Fehlervarianz berücksichtigt wird.

Zunächst zeigt die Bearbeitung der ersten Forschungsfrage, dass die Werte aus den ersten beiden Artikeln in der gleichen Größenordnung liegen. Die ZB für Walderholung liegt mit ca.

27€ im niedrigeren aber trotzdem eindeutig positiven Bereich der ZB für einzelne Waldmerkmale aus Artikel zwei. Wendet man die in Artikeln drei und vier identifizierten Kriterien für Forschungsfrage zwei auf die Ergebnisse des DCE aus Artikel zwei an, sieht man, dass Nicht-Berücksichtigung nicht zu erwarten ist: Nur eine hohe Anzahl von Auswahlkarten hat Auswirkungen auf die Nicht-Berücksichtigung von Merkmalen. Die Zahl der Auswahlkarten lag bei neun Karten, und die getesteten Zahlen lagen bei bis zu 24. Auch die Anzahl von Alternativen liegt mit drei im niedrigen Bereich. Die Fehlervarianz wird zwar von den Designdimensionen beeinflusst, aber die angegebene ZB ändert sich dadurch nicht, wie Artikel vier ergab.

Zur die Beantwortung der dritten Forschungsfrage liegen nun erste Möglichkeiten zur Lösung von Marktversagen vor. Die Möglichkeit zur Einbeziehung öffentlicher Güter in Abwägungsentscheidungen ist nun gegeben und Maßnahmen im Wald können bei Bedarf an den Bedürfnissen der Bevölkerung orientiert werden. Auch für Güter, die bisher nicht bewertet wurden, liegen nun Möglichkeiten vor, ihre Belastbarkeit (in Bezug auf die drei hier betrachteten Kriterien) einzuschätzen.

## **Abbildungs- und Tabellenverzeichnis**

Abbildung 1: Beispiel einer Auswahlkarte mit drei Alternativen und sechs Merkmalen.....	9
Abbildung 2: Unterteilung des Naturschutzindikators .....	24
Tabelle 1: „Design of Designs“ .....	26

# **1. Einführung in die wissenschaftliche Fragestellung**

## **Einleitung**

„Wald im Sinne dieses Gesetzes ist jede mit Forstpflanzen bestockte Grundfläche.“ (BWaldG, §2) Der Wald nimmt in Deutschland ca. ein Drittel der Landesfläche ein und bietet der Bevölkerung vielfältige Ressourcen, die von wirtschaftlichen oder privaten Akteuren genutzt werden. Dabei gibt es aktuell etwas mehr Nadel- als Laubbäume im Wald in Deutschland (ca. 55% Nadelbäume / 45% Laubbäume) und das Durchschnittsalter der Bäume im Wald liegt bei 77 Jahren. Waldmerkmale, die als Lebensräume für vielfältige Tierarten und für die Biodiversität von zahlreichen Pflanzenarten entscheidend sind, wie beispielsweise Totholz und Habitatbäume, sowie die naturnahe Baumartenzusammensetzung, nehmen stetig zu. In Deutschland gilt das Waldgesetz, welches den ersatzlosen Abbau von Waldfläche verbietet. Dies verhindert de facto eine Reduktion der Waldfläche und hat über die Jahre zu einer Zunahme der Waldfläche geführt. Besonders relevant für die Bevölkerung ist die Möglichkeit, den Wald zur Erholung zu nutzen. Das Waldgesetz garantiert der Bevölkerung das Recht, den Wald zur Erholung zu betreten, dies auch unabhängig von der Eigentümerstruktur (BMELV, 2011; BMEL, 2014; BWaldG, §14).

Doch Konkurrenz und Auseinandersetzungen über Ziele bei der Waldgestaltung, den Nutzungsformen oder (waldbaulichen) Maßnahmen führen zu Konflikten zwischen den beteiligten Akteuren. Dies kann die Erholungsleistung des Waldes oder aber politisch gewünschte Leistungen wie Naturschutz oder Artenschutz betreffen. Auch die Forstwirtschaft mit der Holz-Produktionsleistung befindet sich in diesem Spannungsfeld. Darüber hinaus fragt die Bevölkerung neben Erholung und freiem Betreten auch Naturschutz und bestimmte Eigenschaften und Aussehen des Waldes nach. Beeinflusst wird dies durch Gesetzgebungs- und Verwaltungsorgane in Bundesrepublik und Bundesländern, die nationale politische Strategien und Gesetze erlassen, um (internationale und) langfristige Ziele zu verfolgen, sowie um die Angebote, die aus dem Wald kommen, und Nachfrage und Anforderungen an den Wald, zu steuern.

Zur Lösung der Konflikte müssen die Entscheidungsträger in Politik und Wirtschaft Kosten und erwartete Nutzen einzelner Aspekte gegeneinander abwägen, um zu Entscheidungen über die zukünftige Waldbehandlung und entsprechend resultierende Güter zu kommen. Da manche der betroffenen Güter solche sind, die nicht auf Märkten gehandelt werden, gibt es keine vorliegenden Marktpreise, um sie in die Abwägung einzubeziehen. Die Eigenschaften der öffentlichen Güter des Waldes verhindern, dass einzelne Personen von ihrer Nutzung

ausgeschlossen werden können (sie müssen sie nutzen, sie können sie aber auch nutzen wenn sie nichts dafür zahlen). Ein Beispiel könnte an dieser Stelle ein Wald mit Wasserproduktionsleistung sein, von deren nebenbei entstehendem Nutzen in Deutschland alle profitieren, aber niemand ausgeschlossen werden kann. Gleichzeitig gibt es aber auch für den Waldeigentümer keine Möglichkeit, sich die Produktion solch eines Gutes entgelten zu lassen. Zwischen den Nutzern stellt sich bei der Nutzung der öffentlichen Güter keine Rivalität ein, anders als bei einem gefällten Baum, der nur zugunsten des Eigentümers verarbeitet oder verkauft werden kann. Das Vorhandensein von Gütern, die nicht einfach bepreist oder kompensiert werden, führt zu einem Versagen des Marktes mit seinen klassischen Mechanismen zur effizienten Verteilung von Ressourcen, da diese Auswirkungen anderer wirtschaftlicher Aktivitäten, die ohnehin entstehen und nicht kompensiert werden, nur bedingt von Entscheidungsträgern in ihre Abwägungen einbezogen werden können.

Zur möglichen oder erleichterten Einbeziehung der nicht marktlich bewerteten Güter ist das Interesse an verlässlichen Bewertungen aller durch den Wald produzierten Güter in Geldeinheiten groß. Die zur Verfügung stehenden Bewertungsmethoden imitieren Märkte, in dem sie Personen in Experimenten Abwägungsentscheidungen fällen lassen, die nun doch Elemente von Ausschließbarkeit und Rivalität enthalten. So können auch öffentliche Güter mit einer Annäherung an ihren jeweils entstehenden Nutzen bewertet werden. Gleichzeitig sind aber auch diese Bewertungen nicht frei von Bedenken, die sich auf die Herkunft der Daten aus Befragungen und die darauf angewandten statistischen Methoden beziehen. So ist zunächst die Qualität der Befragungsdaten und der Ergebnisse nicht genau bekannt. Auch die gegebenenfalls erforderlichen Vorsichtsmaßnahmen, unter denen die Ergebnisse letztendlich durch die Entscheidungsträger verwendet werden können, sind nicht genau bekannt.

### **Theoretischer Hintergrund und Methode**

Die Ausführungen zum theoretischen Hintergrund dieser Arbeit beginnen bei der Charakterisierung externer Effekte. Externe Effekte entstehen als Nebeneffekte anderer wirtschaftlicher Aktivitäten und werden regelmäßig nicht auf normalen Märkten gehandelt und kompensiert, da die entsprechenden Märkte nicht existieren. Das bedeutet, dass Begünstigte durch diese externen Effekte nichts bezahlen müssen und Benachteiligte nicht entschädigt werden. Dies mag zum Beispiel die Produktion von Erholungsmöglichkeiten für die Bevölkerung sein, wenn die wirtschaftliche Aktivität eigentlich auf die Holzproduktion fokussiert. Die Kräfte des Marktes können in solch einer Situation nicht wirken und Marktversagen entsteht, da Güter vorliegen, die Dritte (positiv oder negativ) beeinflussen, die

aber nicht auf Märkten gehandelt werden. Dies verhindert die pareto-effiziente Allokation von Ressourcen zwischen der Holzproduktion und der Bereitstellung von Erholungsmöglichkeiten, bei der die Verteilung von Ressourcen zur Güter- oder Nutzenproduktion so effizient ist, dass sie nicht mehr verbessert werden kann, ohne einen Akteur schlechter zu stellen (Varian, 2004; Pindyck und Rubinfeld, 2005). Gleichzeitig wird zwischen öffentlichen und privaten Gütern unterschieden (Samuelson, 1954). Die Eigenschaften Nicht-Ausschließbarkeit von der Nutzung und Nicht-Rivalität im Konsum charakterisieren dabei öffentliche Güter, während private Güter beide Eigenschaften in ihrer positiven Ausprägung besitzen. Nicht-Ausschließbarkeit bezieht sich nicht nur darauf, dass Personen nicht vom Konsum ausgeschlossen werden können, sondern auch darauf, dass niemand freiwillig auf die Nutzung verzichten kann, da diese Güter immer allen Personen bereitgestellt werden und niemand sich entziehen kann (Varian, 2004). Dabei existieren für öffentliche Güter keine normalen Märkte, auf denen die Gesellschaft ihre Präferenzen zum Ausdruck bringen könnte. So entstehen Probleme bei der Allokation von Ressourcen, über die die Gesellschaft sich einigen muss, um keinen Wohlfahrtsverlust zu erleiden, weil das Gut gar nicht bereitgestellt wird, oder das Gut, auf Kosten anderer Güter, im Überfluss produziert wird.

Die Arten der ökonomischen Werte und Güter sind vielgestaltig. Sie lassen sich aufteilen in nutzungsabhängige und nutzungsunabhängige Werte. Die eine Gruppe stiftet durch direkten oder indirekten Gebrauch Nutzen, während die andere Gruppe den Nutzen stiftet, indem man die Option, etwas nutzen zu können, sowie die Aussicht, dass folgende Generationen und andere Personen Nutzen haben werden, wertschätzt, ebenso wie das Wissen darüber, dass das Gut überhaupt existiert (Weisbrod, 1964; Krutilla, 1967; Pearce, 2001; Bateman et al., 2002; Bergen et al., 2013). Viele betrachtete Werte in dieser Arbeit haben auch eine starke nutzungsunabhängige Komponente, da die nutzungsabhängigen Werte überwiegend auf Märkten erfasst werden können, und somit die Notwendigkeit anderweitiger Bewertungsmöglichkeiten entfällt.

Durch die Bewertung von öffentlichen Gütern durch Methoden der geäußerten Präferenzen (englisch: stated preferences, SP), die im Umweltbewertungskontext üblich sind, kann ein Markt simuliert und können öffentliche Güter mit Marktmethoden bewertet werden. Resultierend können sie in ökonomische Entscheidungen, beispielsweise über die Bereitstellung von Gütern, einbezogen werden. So kann Marktversagen überwunden werden. Der Nutzen der Bevölkerung eignet sich deshalb für die Bewertung, da er ein Maß für die

Präferenzen der Bevölkerung oder konkreter der Konsumenten der öffentlichen Güter darstellt, und somit eine Rangordnung von Gütern erlaubt. Das an der Stelle entstehende Problem ist der Übergang von ordinalen Nutzenmaßen zu kardinalen Nutzendifferenzen, die eine Aussage darüber erlauben, wieviel besser oder schlechter ein Gut oder ein Güterbündel ist (Varian, 2004; Endres, 2007; Bergen et al., 2013). Dies wird üblicherweise durch die Verwendung von kardinalen Zahlungsbereitschaften überwunden.

Die Entscheidungen über Tausch und Kombinationen von verschiedenen Gütern führen in der neoklassischen Theorie, gemäß der Wohlfahrtsökonomie, zu einer Erhöhung der Wohlfahrt jedes Individuums. Die entstehenden Nutzen aus dem Konsum von (öffentlichen) Gütern und Güterbündeln werden in ordinalen Nutzenfunktionen erfasst. So wird beziffert, welches Gut wieviel Nutzen stiftet, aber nicht wie viel mehr es im Vergleich zu einem anderen stiftet. Die entstehenden Nutzen werden als (gesellschaftliche) Wohlfahrtsfunktionen über alle Individuen zusammengefasst, um letztendlich die Verteilung des Nutzens auf die gesamte Gesellschaft abbilden oder Entscheidungen darüber fällen zu können (Varian, 2004; Pindyck und Rubinfeld, 2005). So basiert die Bewertung von öffentlichen Gütern auf einer (kardinalen) rechnerischen Annäherung an den entstandenen (ordinalen) Nutzen der gesamten Gesellschaft, für welche das Vorgehen im Folgenden erläutert wird.

Zur Lösung der oben genannten Konflikte und zur Überwindung von Marktversagen sind auch für öffentliche Güter verlässliche Bewertungen gefordert. Dies geschieht mittels Methoden, die Nutznäherungen auf anderen Wegen erreichen. Hier werden die Methoden der geäußerten Präferenzen verwendet, die es ermöglichen, trotz der eventuell gegebenen Nutzungsunabhängigkeit, aus den Aussagen von befragten Personen (insbesondere im umweltökonomischen Bereich ist dies häufig die allgemeine Bevölkerung) Präferenzen abzuleiten, indem Märkte imitiert werden (Bateman et al., 2002). Dabei werden Zusammenhänge wie auf klassischen Märkten für private Güter konstruiert, auf denen, unter Beachtung von Budgetrestriktionen, zwischen Gütern und ihren Eigenschaften abgewogen und letztendlich eine Wahl getroffen wird. So können auch öffentliche sowie nutzungsunabhängige Güter, für die kein klassischer Markt existiert, später mit ihren Bewertungsergebnissen in Konkurrenzsituationen positioniert und in Entscheidungen einbezogen werden. So kann das Marktversagen behoben werden, da die Verteilung von Ressourcen abgestimmt auf den entstehenden Nutzen erfolgen kann.

Methoden der geäußerten Präferenzen erfassen in Befragungen direkt die Präferenzen und Zahlungsbereitschaften (ZB), und sind somit als Bewertungsmaß für Güter geeignet, bei denen häufig keine beobachtbaren Präferenzen vorliegen. Um die Nutzennäherung zu erreichen, wird den Befragten die Möglichkeit zur Präferenzäußerung gegeben. Dazu werden sie gebeten, ihre Präferenzen durch Wahlen und Angaben von ZB zu nennen und darauf, nicht auf ihren tatsächlichen Handlungen, basiert die Bewertung. So lässt sich der entstehende Nutzen näherungsweise beziffern. Mittels dieser Nutzennäherung können nicht-marktliche Güter in (politische) Abwägungsentscheidungen zwischen Leistungen des Waldes einbezogen werden.

Zwei in der umweltökonomischen Bewertung besonders verbreitete Methoden der geäußerten Präferenzen sind die bedingte Bewertung (englisch: Contingent Valuation Method, CVM) und diskrete Auswahlexperimente (englisch: Discrete Choice Experiments, DCE). Methoden der geäußerten Präferenzen versetzen die Befragten über den experimentellen Versuchsaufbau in eine Situation, in der sie ähnlich wie auf einem Markt Entscheidungen fällen sollen. Dabei stehen ihnen Informationen zur Verfügung, es werden Restriktionen genannt und es gibt die Möglichkeit, Präferenzen über Güter zu äußern, indem man sie auswählt, in eine Präferenzordnung bringt oder eine ZB äußert. Die Fragetechniken unterscheiden sich zwischen den Methoden, ebenso unterscheidet sich auch die resultierende Struktur der erhobenen Daten. Bei der Verwendung von CVM werden die Befragten häufig direkt nach ihrer selbst geäußerten ZB für ein klar umrissenes Gut gefragt. So lassen sich später eine Verteilung der ZB-Werte und eine entsprechende Nachfragekurve abbilden. In DCE werden Auswahlentscheidungen von den Befragten erbeten, über die sie ihre Präferenzen über mögliche Zahlungen zum Ausdruck bringen. Die resultierenden Daten erlauben eine Wahrscheinlichkeitsaussage über die Wahl einer Option, charakterisiert durch unterschiedliche Eigenschaften, sowohl über inhaltliche Merkmale als auch über monetäre Merkmale die als ZB interpretiert werden können.

Die Verwendung von CVM bedeutet die Erstellung einer hypothetischen Marktsituation, in der ein präzise beschriebenes Gut in einer bestimmten Menge angeboten wird und die Befragten nach ihrer maximalen ZB unter den beschriebenen Umständen gefragt werden. Dies soll die Entscheidung, die auf einem Markt getroffen wird, imitieren. In der ursprünglichen Definition nach Mitchell und Carson (2005) bezieht sich diese Methode auf die Bewertung öffentlicher Güter mittels einer bestimmten Befragungstechnik. So erlaubt die

CVM Aussagen über die Wohlfahrtswirkung öffentlicher Güter (Carson und Louviere, 2011). Grundsätzlich ergibt sich aus der Frage eine Verteilung von Antworten, aus denen Verteilungsmaße der ZB errechnet werden können. Desweiteren kann die Nachfragekurve geschätzt und abgebildet werden. Auch eine Hochrechnung auf die Bevölkerung ist so möglich. Für die Bewertung ergibt sich die über alle Personen aggregierte Konsumentenrente. Unterschiede in der Durchführung von CVM zeigen sich in der Methode der Kontaktaufnahme mit den Befragten (bspw. persönlich oder schriftlich; Lindhjem und Navrud, 2011), in der Menge und Präzision der zur Verfügung gestellten Informationen über die Zahlungsmodalitäten (bspw. die Budgetbeschränkung; Bateman et al., 2002) und schließlich über das Frageformat (bspw. als Auktion oder „take it or leave it“-Verfahren; Mitchell und Carson, 2005; Bergen et al., 2013).

CVM wurden erstmals 1963 in den USA zur Bewertung von Möglichkeiten zur Jagd verwendet (Davis, 1963). Einem viel größeren Personenkreis wurde die Methode bekannt, als die National Oceanic and Atmospheric Administration (NOAA; Arrow, 1993) nach der Exxon Valdez-Ölkatastrophe 1989 CVM für die Erfassung von nutzungsunabhängigen Werten der Bevölkerung verwendete. Seitdem wurde sie in zahlreichen Studien in Deutschland und in anderen Ländern für die Bewertung von unterschiedlichen nicht-marktlichen Gütern verwendet.

Unumgänglich für die Bewertung mittels DCE ist die Annahme, dass Güter in ihre Eigenschaften teilbar sind. Eigentlich stiftet erst die Kombination der Eigenschaften (und auch die Kombination mit den Eigenschaften anderer Güter) Nutzen, und nicht das Gut per se (Lancaster, 1966; Rosen, 1974). Dadurch stellt sich nicht nur die Frage, welches Gut gewählt wird (im vorliegenden Fall bedeutete dies, welche Kombination von Eigenschaften im Wald bevorzugt würde), sondern auch, durch welche Eigenschaften des Gutes die Nachfrage nach dem Gut hervorgerufen wird, und in welchem Zusammenhang dies zur ZB steht. Diese Annahme legt einen theoretischen Grundstein für die Verwendung von DCE, bei der verschiedene Eigenschaften von Gütern einzeln einfließen und bewertet werden können.

Die Verwendung des Ansatzes der Teilbarkeit von Gütern in Eigenschaften führt zur Zufallsnutzentheorie (englisch: random utility theory, RUT), die von Nutzenmaximierung unter dem Einfluss von beobachtbaren oder systematischen und zufälligen Komponenten ausgeht (Thurstone, 1927; Manski, 1977; Louviere et al., 2006; Train, 2009)

*Nutzen: U; systematische Komponente im Nutzen: V; zufällige Komponente im Nutzen: ε*  
*Person: n; Alternativen: i, j; Wahrscheinlichkeit: P; Auswahlkarte: C*

$$U_{ni} = V_{ni} + \varepsilon_{ni} \quad (1)$$

wobei sich die systematische Komponente aus den Einflüssen der einzelnen Eigenschaften  $X - Z$  des Gutes zusammensetzt und eine Eigenschaft kann der Preis sein

$$V_{ni} = \beta_X X_{ni} + \dots + \beta_Z Z_{ni} \quad (2)$$

Die Wahrscheinlichkeit für die Wahl einer Alternative  $i$  durch Person  $n$  lautet

$$P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \text{ für alle } i, j \in C, j \neq i \quad (3)$$

und hängt von den erzielten Nutzen der Alternativen ab. Unter Verwendung von diskreten Auswahlmodellen und beispielsweise dem „conditional logit model“ (McFadden, 1974; Louviere und Woodworth, 1983) kann der Einfluss der einzelnen Merkmale auf den Gesamtnutzen bestimmt werden. So werden die Parameter der einzelnen Merkmale ( $\beta$ ) und Auswahlwahrscheinlichkeiten der Alternativen oder Merkmalsbündel quantifiziert. Gleichzeitig können die Einflüsse der einzelnen Merkmale auf den Nutzen und Austauschbeziehungen zwischen den Merkmalen betrachtet werden (Louviere et al., 2006; Train, 2009). Aus den Austauschbeziehungen zwischen dem Parameter des Preismerkmals und den Parametern der anderen Merkmale lassen sich Zahlungsbereitschaften errechnen. Unter der Annahme, dass eine der angebotenen Alternativen für die Befragten einfach (ohne Zahlung) realisierbar ist, sind die Werte wohlfahrtstheoretisch korrekt (Hanley et al., 2001). Dies wird üblicherweise durch die Verwendung einer Status Quo-Alternative umgesetzt. Diese gewährleistet, dass die Befragten zwar gezwungen sind, eine Wahl zu treffen, diese aber auf keine der hypothetischen Alternativen fallen muss, sondern auch auf das Beibehalten des Status Quo fallen kann. Bei Fehlen einer Status Quo-Alternative sind die Befragten gezwungen eine Wahl zwischen zwei hypothetischen Alternativen zu treffen, die fälschlich als Nachfrage für eine der Alternativen interpretiert wird.

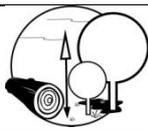
DCE setzen sich bei der praktischen Umsetzung üblicherweise aus mehreren Elementen zusammen. Die Dimensionalität bezieht sich auf die Größe und Komplexität des DCE. Diese

entsteht durch die Kombination von Zeilen und Spalten, die ein DCE ausmachen und die Abwägung und die Auswahlentscheidung verkomplizieren können, wenn die Befragten abwägen und ihre bevorzugte Alternative auswählen (Louviere et al., 2006). Bei einem DCE werden fünf Dimensionen unterschieden:

1. die Anzahl der unterschiedlichen Merkmale, die in Zeilen geschrieben werden, und über die Merkmalsausprägungen gegeneinander abgewogen werden müssen
2. die Anzahl an möglichen Alternativen (Merkmalskombinationen), teilweise inklusive einer Status Quo-Alternative, teilweise ohne. Die Entscheidung für eine der Alternativen ist das Resultat der Auswahlentscheidung.
3. die Anzahl der Auswahlkarten, auf denen nacheinander Entscheidungen gefällt werden müssen,
4. und 5. die Merkmalsausprägungen (die eigentlich die Merkmale erst charakterisieren, da diese in den Zeilen gleich sind, sich nur in den Zeilen durch die Ausprägungen unterscheiden) mit ihrer Anzahl an möglichen Levels pro Merkmal und der Bandbreite über die sie sich vom untersten bis zum obersten Level erstrecken.

Abbildung 1 zeigt ein Beispiel für eine Auswahlkarte mit sechs Merkmalen und drei Alternativen.

**Wenn nur die folgenden Möglichkeiten für die Entwicklung der Landschaft im Umkreis von bis zu 15 Kilometern um den Ort, an dem Sie wohnen, zur Verfügung stünden, welche würden Sie wählen?**  
**Wenn Sie in einer großen Stadt wohnen, denken Sie bitte an die Umgebung dieser Stadt.**

	<b>Landschaft A</b>	<b>Landschaft B</b>	<b>Landschaft C</b>
 <b>Anteil Wald an der Landschaft</b>	10% niedriger	10% höher	Wie heute
 <b>Größe einzelner Felder und Waldstücke</b>	Halb so groß	Doppelt so groß	Wie heute
 <b>Waldflächen mit Unterwuchs</b>	Halb so viele	Doppelt so viele	Wie heute
 <b>Nadelbaumanteil</b>	30% Nadelbaumanteil	70% Nadelbaumanteil	Wie heute
 <b>Erntealter der Wälder</b>	20 Jahre später ernten	30 Jahre später ernten	Wie heute
 <b>Finanzieller Jahresbeitrag zum Landschaftsfonds</b>	25 €	80 €	0 €
<b>Ich wähle</b>			

**Abbildung 1: Beispiel einer Auswahlkarte mit drei Alternativen und sechs Merkmalen** (verwendet nach Weller und Elsasser, 2018; hier in der Originalformulierung)

Kritikpunkte an SP beziehen sich auf die Ausprägungen CVM und DCE gleichermaßen. Dies bezieht sich zunächst auf die hypothetische Situation, in der die Befragten ihre Antworten geben und ihre Zahlungsbereitschaften so angeben können, dass sie ihnen einen „warm glow“ bescheren, also die moralische Befriedigung, Gutes getan zu haben (Kahneman und Knetsch, 1992). Zahlreiche Studien beziehen sich darüber hinaus auf die Anforderungen an das Studiendesign von SP-Studien (Bateman et al., 2002), Frageformate und Durchführung und legen Kriterien an, die die Validität der verwendeten Methode beurteilen (Kling et al., 2012; Johnston et al., 2017). Ein weiterer Kritikpunkt ist der verwendete Mengenbezug des Gutes (englisch: embedding; bspw. Bewertung einer einzelnen Tierart oder aller Tierarten eines Lebensraums) in der Bewertung, da die resultierenden Werte häufig fast gleich hoch sind,

obwohl die Menge sehr unterschiedlich ist (Kahneman und Knetsch, 1992; Hausman, 2012). Abhängig von der Fragestellung und -formulierung zielen die Bewertungen mehr oder weniger auch auf die Erfassung von nutzungsunabhängigen Werten ab.

Vor- und Nachteile der einen oder anderen Methode lassen sich eher in der Interpretation der geäußerten ZB und Präferenzen finden. CVM erlauben eher eine Gesamtbewertung eines speziellen Gutes oder eines Programms, während DCE differenziertere Ergebnisse für einzelne Eigenschaften eines Gutes liefern (Adamowicz et al., 1998). Studien, die die erhobenen ZB-Ergebnisse der beiden Methoden vergleichen, fanden, dass die Ergebnisse stark von der Frageformulierung und den anschließenden Spezifikationen der Berechnungen abhängen, aber die Werte sich häufig nicht statistisch unterscheiden (Adamowicz et al., 1998; Mogas et al., 2006).

Inzwischen gibt es Diskussionsbeiträge, die sich gegen die weitere Verwendung von CVM aussprechen, da die Ergebnisse aus den oben angesprochenen Gründen zu unsicher und die Aussagekraft limitiert ist (Hausman, 2012; Desvousges et al., 2016). Andererseits sprechen sich andere für weitere Untersuchungen von CVM aus, unter anderem um die Anreize für Befragte, ihre Präferenzen zu äußern, zu verbessern (Haab et al., 2013).

### **Überblick über Literatur und forstpolitische Programme**

Die Werte öffentlicher Güter sind vor allem in politischen Kontexten hilfreich, wenn Maßnahmen zugunsten unterschiedlicher Ziele ergriffen und in Gesetzen umgesetzt werden sollen, wo bisher Werte oder Informationen nicht oder nur lückenhaft vorhanden sind. An der Stelle können Bewertungsergebnisse idealerweise eingesetzt werden. Es existieren zahlreiche politische Programme und wissenschaftliche Veröffentlichungen die sich zentral oder am Rande mit der Behandlung des Waldes beschäftigen und die Wichtigkeit öffentlicher Güter betonen.

Die Nationale Strategie zur biologischen Vielfalt (Biodiversitätsstrategie; BMU, 2007) setzt das UN-Übereinkommen über die biologische Vielfalt um. Für den Wald lassen sich als Ziele mehr naturnahe Waldflächen, mehr heimische Baumarten und ein hoher Naturschutz- und Erholungsfaktor, sowie möglichst große unzerschnittene Waldflächen ableiten. Zur Überprüfung der deutschen Nachhaltigkeitsstrategie werden in regelmäßigen Abständen zahlreiche Indikatoren ausgewertet, die auf unterschiedlichen Ebenen nachhaltiges Wirtschaften und Leben in Deutschland messen. Darunter befindet sich der Indikator

„Artenvielfalt und Landschaftsqualität“ mit dem Ziel, gesündere Lebensräume zu erreichen (Destatis, 2016). Der Teilindikator für Wälder (2013: 86,9% des Zielwertes 100) zeigt einen höheren Wert als die Teilindikatoren für Agrarlandschaften, Siedlungen und Gewässer. Auch die Waldstrategie 2020 formuliert als Ziele mehr Biodiversität im Wald und mehr heimische Baumarten bei weiterhin großem Erholungsangebot (BMELV, 2011).

In diesen Strategien sind zunächst die Ziele formuliert, allerdings immer (insb. in der Waldstrategie) im Spannungsfeld zu anderen Waldnutzungsformen. Insbesondere bei der Betrachtung der Holzproduktion liegt es auf der Hand, dass zumindest entgangenen Gewinnen etwas gegengerechnet werden muss, will man andere Ziele als die Holzproduktion verfolgen.

Zunächst sei kurz auf den Hintergrund eingegangen, der sich mit Studien zum Wert der Natur für den Menschen beschäftigt. Die sogenannten Ökosystemleistungen sind Leistungen, die die Ökosysteme für die Menschen erbringen. Eigentlich hat die Natur für die Menschen einen unschätzbaren Wert, dies in einer Bewertung darzustellen überstiege aber ein sinnvolles Maß (Costanza et al., 1997) und muss deshalb anders, und in Bezug auf konkrete Fragestellungen, erfasst werden. 2005 gibt das MEA einen ersten systematischen Überblick über zahlreiche konkrete Ökosystemleistungen unter Verwendung bestimmter Definitionen (Daily, 1997). Die Ergebnisse zeigen große Schäden an den Ökosystemleistungen der Erde, zeigen aber auch Wege auf, das weitere Fortschreiten zu verhindern (MEA, 2005). Dazu bedarf es aber der monetären Bewertungen der Leistungen. Zur Bewertung von Ökosystemleistungen und als Gesamtschau der vorhandenen Bewertungsansätze für Ökosystemleistungen stehen die TEEB-Studien zur Verfügung, die auch in Deutschland umgesetzt werden (u.a. Bösch et al., 2018).

Im Hinblick auf Ergebnisse der umweltökonomischen Forschung sei zunächst eine Überblicksstudie genannt, die Studien in Deutschland, Österreich und der Schweiz zu Umweltbewertung sammelt und einen Überblick gibt (Elsasser et al., 2016). Hier zeigt sich, dass es vor allem für Biodiversität im weitesten Sinne (u.a. Meyerhoff et al., 2012, CVM zur Umsetzung einer Biodiversitätsstrategie im Wald) und für Erholung zahlreiche Studien gibt. Speziell Erholung mag für die Befragten ein einfacher zu bewertendes Gut sein (Binder et al., 2017), da es stärker nutzungsabhängig ist und die Befragten konkretere Vorstellungen davon haben, als bspw. von Kohlenstoff-Sequestrierung. Aber auch zu anderen Gütern gibt es Fallstudien (Olschewski et al., 2012, DCE zu Lawinenschutz in der Schweiz; Ryffel et al.,

2014, DCE zu Hochwasserschutz in der Schweiz). Speziell in der Verhandlung mit Waldbesitzern und Forstbetrieben wären Werte für die Kompensation von erbrachten öffentlichen Gütern (wie bspw. Kohlenstoff-Sequestrierung) von Nutzen, hier wurde ein Bedarf formuliert (Möhring und Mestemacher, 2009).

Bei der Bewertung der Erholungsleistung geht es insbesondere darum, den monetären Erholungswert der Wälder für Alltagsbesucher festzustellen. Dies war in der Vergangenheit zunächst für den jeweils aktuellen Zustand des Waldes geschehen, ohne mögliche Veränderungen der Wälder in die Bewertung einzubeziehen (Klein und Elsasser, 1994; Löwenstein, 1994; Bergen et al., 1995, Elsasser, 1996).

Für die Bewertung von Landschaft und Naturschutz wird das zu bewertende Gut für eine bessere Konkretisierung häufig in mehrere einzelne Merkmale unterteilt, die dann letztendlich einzeln untersucht und bewertet werden können. Zum einen wird es so greifbarer für die Befragten, zum anderen ergibt sich so auch die Möglichkeit, mehrere Aspekte von Landschaft und Naturschutz zu betrachten und einzeln monetär zu bewerten (Sagebiel et al., 2017). So können die Einzelwerte im Zusammenhang mit (nationalen) politischen Strategien betrachtet werden. Bisher war die Bewertung häufig im Zusammenhang mit anderen Gütern und als Teil von Programmen erfolgt. Die bisher daraus hervorgehenden Resultate deuten darauf hin, dass Laubwälder und Diversität geschätzt werden (Elsasser et al., 2010; Abildtrup et al., 2013; Giergiczny et al., 2015; Filyushkina et al., 2017).

Vergangene, eher methodische, Studien haben gezeigt, dass zusätzlich zu eventuell bereits vorhandenen unterschiedlichen bewussten oder unbewussten Mechanismen (Antwortheuristiken), Entscheidungen und geäußerte ZB auch durch die Charakteristika des DCE beeinflusst werden (Scarpa et al., 2009; Hess und Hensher, 2010; Campbell et al., 2011; Hensher et al., 2012). Dies bedeutet, dass diese Komplexität auf die Antworten der Befragten einwirkt, da die Fähigkeit von Menschen, komplexe Informationen zu verarbeiten, generell begrenzt ist (Swait und Adamowicz, 2001). Wissenschaftler geraten in Versuchung, die Komplexität eines DCE durch die Aufnahme weiterer Auswahlsets oder Merkmale zu erhöhen, um mehr Beobachtungen oder Informationen zu mehr Merkmalen zu erhalten (Hensher, 2004; Caussade et al., 2005; Hensher et al., 2005). Dies geschieht unter größerer kognitiver Inanspruchnahme der Befragten und führt möglicherweise zu verzerrten Befragungsergebnissen aufgrund der stärkeren Inanspruchnahme (DeShazo und Fermo, 2002). Dies kann unter anderem die resultierende Nicht-Berücksichtigung von einzelnen

Merkmale betreffen, aber ebenso die Abbruchquote der Befragung und die Fehlervarianz der Antworten der Befragten oder die Häufigkeit der Wahl der Status Quo-Alternative.

Speziell bei der Untersuchung der Nicht-Berücksichtigung von Merkmalen in DCE können die Aussagen der Befragten einerseits (stated attribute non-attendance), und ihr gezeigtes Verhalten in der Befragung andererseits (inferred attribute non-attendance), differieren. In der Literatur wird dieser Unterschied häufig durch Kombination dieser beiden Informationen berücksichtigt (Hensher et al., 2007; Hensher und Rose, 2009; Campbell et al., 2011; Kragt, 2013; Scarpa et al., 2013). Bisher wurde nur von Hensher (2006) systematisch der Zusammenhang zwischen Nicht-Berücksichtigung und den Designdimensionen untersucht und ein Einfluss aller Designdimensionen auf die Nicht-Berücksichtigung von Merkmalen gefunden. Ein Kritikpunkt der Untersuchung von Nicht-Berücksichtigung von Merkmalen ist der Hinweis, dass es sich dabei nicht notwendigerweise um die Überforderung der Befragten handeln muss, sondern auch einfach ein Ausdruck unterschiedlicher Präferenzen sein kann, der dazu führt, dass Merkmale nicht berücksichtigt werden (Balcombe et al., 2011; Hess et al., 2013). Die Abbruchquote wird von einer größeren Belastung der Befragten durch die Komplexität des DCE beeinflusst. So finden Hensher et al. (2001) einen schwachen Effekt, einen negativen Einfluss findet Galesic (2006), Bech et al. (2011) finden keinen signifikanten Effekt, und ein schwacher Effekt wird gefunden bei Louviere et al. (2013). Eine größere Anzahl von Auswahlentscheidungen nacheinander kann nach einem Lerneffekt aber auch zu Ermüdungserscheinungen und einer entsprechend größeren Fehlervarianz führen, die auf eine größere Entscheidungsunsicherheit hinweist (Dellaert et al., 1999; DeShazo und Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005; Rolfe und Bennett, 2009; Chung et al., 2011 (12 Alternativen); Czajkowski et al., 2014 (26 Auswahlkarten)). Ebenso ist der Effekt von mehr Alternativen, die gleichzeitig zu betrachten sind, unklar, es gibt aber Hinweise in der Literatur (Arentze et al., 2003; Caussade et al., 2005; Chung et al., 2011), dass bei mehr Alternativen eher eine dabei ist, die den Präferenzen des Befragten gut entspricht, und die Auswahl leicht und mit größerer Sicherheit getroffen werden kann (DeShazo und Fermo, 2002). Die Hinweise dafür, dass eine größere Anzahl von Merkmalen die Entscheidung für eine Alternative immer schwieriger macht, sind eindeutig (DeShazo und Fermo, 2002; Arentze et al., 2003; Caussade et al., 2005).

## **Forschungsfragen**

Eine Lösung für die aus Theorie und Literatur abgeleiteten Probleme stellt die eindeutige Bewertung aller Güter dar. Wie bereits erläutert, kann der Markt allein diese Funktion für die

öffentlichen Güter nicht erfüllen. Aus diesem Grund müssen andere Methoden zur Anwendung kommen, um eine Bewertung vorzunehmen. Darüber hinaus müssen möglicherweise vorhandene Schwächen dieser Bewertungsmethoden aufgedeckt und geprüft werden. Dies führt zu den ersten beiden Zielen dieser Arbeit: zum einen soll für Güter, um die sich Konflikte drehen und die bisher keinen bezifferten (oder keinen aktuellen) Wert haben, einen Wert erhoben und berechnet werden. Zum anderen soll die Methode, die zur Bewertung verwendet wurde, bezüglich ausgewählter Kriterien überprüft werden. Darüber hinaus sollen als drittes Ziel die Werte verwendet werden können, um öffentliche Güter in (politische) Verhandlungen einzubeziehen und so das Marktversagen zu überwinden.

Mit den vorliegenden Artikeln eins und zwei werden Bewertungsergebnisse präsentiert, die in den oben angesprochenen Konflikten zwischen unterschiedlichen Interessensgruppen für eine klarere Verhandlung sorgen können. Die Euro-Werte können in Abwägungsentscheidungen zur Beurteilung entstehender Kosten und in Verhandlungen über die Verteilung von Rechten einbezogen werden.

Gleichzeitig helfen die vorliegenden Artikel drei und vier dabei, die Qualität der verwendeten Erhebungsmethode „geäußerte Präferenzen“ (hauptsächlich DCE in Artikel zwei), in Bezug auf die untersuchten Kriterien, besser einschätzen zu können. In dieser Arbeit werden nun Ergebnisse für diese beiden Ziele innerhalb des Kontexts Wald präsentiert.

Die abgeleiteten Ziele für diese Arbeit in drei Punkten zusammengefasst sind

- die kritische Anwendung der Methoden CVM und DCE und die Ableitung von Bewertungsergebnissen für einzelne Leistungen des Waldes (Erholung und „Landschaft und Naturschutz“)
- die Untersuchung von methodischen Aspekten der DCE, insb. hinsichtlich Nicht-Berücksichtigung von Merkmalen sowie Abbruchquote und Fehlervarianz der Antworten in Abhängigkeit von der Komplexität des Befragungsdesigns
- die Möglichkeit anzubieten, Güter in (politischen) Verhandlungen zu positionieren und so Marktversagen teilweise zu überwinden

## **2. Einordnung der eigenen Ergebnisse in den fachlichen Kontext**

Im Folgenden werden vier Artikel im Kontext der Forschungsfragen dieser Arbeit vorgestellt. Die ersten beiden Artikel benutzen die Methoden CVM und DCE zur Bewertung von öffentlichen Gütern des Waldes durch die Bevölkerung. Damit tragen sie zur Beantwortung der ersten Forschungsfrage bei, die nach Bewertungsergebnissen für einzelne Leistungen des Waldes fragt. Bewertet wird deutschlandweit die Erholungsleistung des Waldes, inklusive einer Abschätzung der möglichen Obergrenze der Zahlungsbereitschaften. Daraus folgen Werte für die Erholungsleistung, sowie eine mögliche Obergrenze für die Höhe der Werte. Im zweiten Artikel wird eine deutschlandweite online-Befragung zum Wert von Naturschutz und Landschaft dargestellt. Dazu wird ein DCE herangezogen. Es werden Zahlungsbereitschaftsergebnisse für verschiedene Merkmale des Waldes präsentiert, die alle auf Landschaft und Naturschutz abzielen und die flächendeckend deutschlandweit bisher nicht vorgelegen haben.

Aus der Anwendung der Methode der Auswahlexperimente und einem aus der Literatur vorliegenden „Design of Designs“-Plan (Hensher, 2004) heraus ergibt sich die Frage, welchen Einfluss die Komplexität des Designs auf die Entscheidung der Befragten hat. Die letzten beiden Artikel untersuchen die Methodik der DCE näher und unter welchen Bedingungen sie valide Ergebnisse bringt. Zentrale Punkte sind dabei die Nicht-Berücksichtigung von einzelnen Merkmalen im Auswahlexperiment, sowie das Abbruchverhalten und die Fehlervarianz der Antworten der Befragten, in Abhängigkeit von der Dimensionalität oder Komplexität eines Auswahlexperiments. Hieran lässt sich erkennen, ob bspw. viele Merkmale auf einmal oder viele Auswahlkarten nacheinander die Antworten verändern. Damit tragen sie zur Beantwortung der zweiten Forschungsfrage bei, die eine methodische Untersuchung von DCE fordert. Die Ergebnisse dieser beiden ersten Forschungsfragen können dann zur Beantwortung der dritten Forschungsfrage verwendet werden, die eine Möglichkeit zur Verwendung der Ergebnisse in politischen Verhandlungen sucht.

Alle verwendeten Datensätze stammen aus allgemeinen Bevölkerungsbefragungen. Dies bedeutet, dass die gesamte Bevölkerung die Zielgruppe darstellt und die Aussagen entsprechend allgemein verwendet werden können ohne nur für Subgruppen zu gelten. In der Diskussion zu Artikel eins wird das noch zu behandeln sein. Die Fragebögen waren jeweils von den Autoren der Studien entwickelt und dann von Befragungsinstituten durchgeführt, die insbesondere für die online-Stichproben ihre jeweiligen Panel-Mitglieder kontaktierten. Dies, aber auch die Ziehung einer Stichprobe aus der Wohnbevölkerung, soll eine repräsentative

Abbildung der Bevölkerung gewährleisten. Speziell bei der ersten Befragung wurde der Fragebogen aus einer vergangenen Studie des einen Autors übernommen und nach kleineren Überarbeitungen verwendet. Grundsätzlich bestehen die Fragebögen aus kurzen Einleitungsfragen zu persönlichen Merkmalen, dann folgt der Fragenteil zur Bewertung des öffentlichen Gutes, anschließend werden eventuell Fragen zur Entscheidungssicherheit o.ä. gestellt und erst danach folgen weitere Fragen zu Einstellungen und Hintergründen. Dies gewährleistet, dass die Befragten bei der Beantwortung der Bewertungsfragen einerseits noch nicht „müde“ sind und andererseits die Antworten nicht durch Hinweise, die möglicherweise in den anderen Frageblöcken gegeben werden, beeinflusst werden.

Der Datensatz zum ersten Artikel beinhaltet 2060 Interviews vom November und Dezember 2011, jeweils etwa die Hälfte als online- und persönliche Interviews umgesetzt. Für die persönlichen Befragungen wurde eine mehrstufige systematisch-randomisierte Stichprobe aus der deutschen Wohnbevölkerung gezogen. Der Datensatz für Artikel zwei beinhaltet 2932 Interviews aus einer online-Befragung im März und April 2013. Auch der dritte Datensatz stammt aus einer online-Befragung vom Dezember 2012 und beinhaltet 1684 Interviews. Dieser wurde für die Auswertungen für die Artikel drei und vier verwendet. Die verwendeten Merkmale in den Datensätzen zwei und drei sind teilweise überlappend, da es sich bei Datensatz drei um eine Vorstudie für die Befragung für Datensatz zwei handelt. Durch den methodischen Untersuchungsanspruch sind die Anforderungen an das verwendete Design, sowie die genaue Ausgestaltung der Merkmale aber andere, wie an entsprechender Stelle noch ausgeführt wird.

## **2.1 Erster Artikel „Aktuelle und potentielle Erholungsleistung der Wälder in Deutschland: Monetärer Nutzen der Erholung im Wald aus Sicht der Bevölkerung“ (Elsasser und Weller, 2013, AFJZ)**

Der Artikel wurde von Dr. Peter Elsasser (Federführung) und Priska Weller auf Deutsch geschrieben und 2013 nach einem peer-review bei der Zeitschrift „Allgemeine Forst- und Jagdzeitung“ veröffentlicht. Die Vorbereitungen für die Befragung wurden gemeinsam durchgeführt, die Berechnungen führte Dr. Peter Elsasser durch, Priska Weller trug zur Diskussion bei.

### **Zusammenfassung**

Um die Erholungsleistung der deutschen Wälder umweltökonomisch zu bewerten, wurde eine Bevölkerungsbefragung in Deutschland durchgeführt. Die ZB für das zu bewertende Gut

wurde mittels CVM erhoben. Das zu bewertende Gut wurde als das Recht, den Wald in der Wohnumgebung ein Jahr lang zur Erholung betreten zu dürfen, beschrieben und die ZB dafür erfragt. Bei der CVM werden die Befragten in einer hypothetischen und beschriebenen Marktsituation direkt nach ihrer maximalen individuellen ZB gefragt, als Ausweichmöglichkeit wird die Nicht-Nutzung des angebotenen Gutes bei Nicht-Zahlung angeboten. Um die Obergrenze der ZB abschätzen zu können, wurde im Anschluss an die Frage nach der ZB nach gewünschten Verbesserungen im Wald und resultierender ZB dafür gefragt.

Zunächst (um die genutzte „Menge“ des Gutes Walderholung einschätzen zu können) wurde die Anzahl der Waldbesuche im vor der Befragung vergangenen Jahr erfragt. Aus den Daten geht hervor, dass gut drei Viertel der Bevölkerung mindestens einmal im Jahr den Wald zur Erholung besuchen. Etwa 80% der Waldbesuche finden im Umkreis von ca. 5 km um den Wohnort statt (es wurde nach der Entfernung zum zuletzt besuchten Wald in der Wohnumgebung gefragt). Als Ergebnis der Untersuchung wird aus den angegebenen jährlichen ZB der Durchschnitt für die Gesamtbevölkerung (ab 14 Jahren, da nicht nur Erwachsene den Wald zur Erholung nutzen und dies die Alters-Untergrenze der Befragten war) errechnet; durch die Betrachtung der unterschiedlichen Befragungsmodi sowie die Beschränkung auf Waldbesucher oder alle Personen, ergeben sich unterschiedliche Subgruppen mit leicht unterschiedlichen ZB. So ergeben sich für die persönlich Befragten ZB von 26,94€/p/a, werden nur die Waldbesucher einbezogen erhöht sich die ZB auf 32,06€/p/a. Die online Befragten haben eine durchschnittliche ZB von 30,32€/p/a, werden auch hier nur die Waldbesucher einbezogen erhöht sich die ZB auf 32,38€/p/a. Aus der zusätzlich durchgeführten Regressionsanalyse wird ersichtlich, dass Einkommen und Besuchshäufigkeit des Waldes große Einflüsse auf die ZB haben, mehr als andere persönliche Eigenschaften der Befragten, die entweder geringer oder gar nicht signifikant auf die jährliche ZB der Befragten einwirken.

Um die Obergrenze der ZB und den potentiellen Erholungswert der Wälder abschätzen zu können, wurde anschließend in mehreren möglichen Antwortkategorien nach gewünschten Änderungen im Wald gefragt. Diese Fragen ergaben, dass über die Hälfte der Befragten sich Änderungen an der Infrastruktur wünscht, ein gutes Viertel der Befragten wünscht sich Änderungen hin zu mehr Naturschutz im Wald und das verbleibende knappe Fünftel hat keine Verbesserungswünsche. Allerdings entstehen auch nach der hypothetischen Realisierung dieser Wünsche kaum zusätzliche Waldbesuche der Befragten, sowie nur eine geringe

zusätzliche ZB i.H.v. ca. 1€/p/a bei den persönlich Befragten und ca. 4€/p/a bei den online Befragten. Auch das Angebot, drei Verbesserungswünsche umzusetzen ergab kein anderes Bild bei den Befragten. Daraus lässt sich die Hypothese ableiten, dass der Nutzen von Walderholung sich kaum steigern lässt.

### **Diskussion und Einordnung in den thematischen Rahmen**

Zur tieferen Untersuchung der Ergebnisse werden verschiedene methodische Details betrachtet. Zunächst stellt sich die Frage nach den Auswirkungen der verwendeten Erhebungsmethode. Insgesamt unterscheiden sich die ZB-Ergebnisse aus den beiden Befragungsmodi nicht signifikant. Nur die Waldbesucheranteile und Besuchshäufigkeiten unterscheiden sich zwischen den Befragungsmodi, sowie der Anteil der Personen, der Einzelfragen nicht beantwortet hat. Auch in der Literatur besteht Konsens darüber, dass online-Befragungen nicht schlechter sind als persönliche Befragungen, da Selbstselektion und Antwortverweigerung der Befragten in allen Befragungen die Ergebnisse beeinflussen (Lindhjem und Navrud, 2011).

Des Weiteren war diese Untersuchung die Aktualisierung einer in den 1990er Jahren durchgeführten Studie, aber mit methodischen Änderungen. In beiden Studien wurde die Methode der CVM verwendet, allerdings anders als in den 1990er Jahren diesmal für eine deutschlandweite Haushaltsstichprobe, nicht für Fallstudien in den Zielgebieten (Waldbesucherbefragung) mit anschließender Hochrechnung auf die Bundesrepublik. Auch die oben bereits thematisierte zusätzliche Verwendung einer online-Stichprobe war hier neu. Die Ergebnisse unterscheiden sich nun in der Höhe der (nominalen) Werte und sind in der neuen Studie niedriger als in der älteren. Die direkte Vergleichbarkeit der Ergebnisse ist durch die methodischen Änderungen in der neuen Durchführung eingeschränkt. Unter anderem könnte der Ort der Umfrage die angegebene ZB beeinflusst haben, da er einmal im Wald und einmal zuhause lag. So war das bewertete Gut weniger präsent. Insbesondere die Wahl der Befragten-Stichprobe und die damit verbundene Verzerrung bei der Hochrechnung auf die Bundesrepublik in den 1990er Jahren und die Standorte der Durchführung der Befragungen könnten als Erklärungen für die niedrigere ZB im Vergleich zur Vorgängerstudie dienen.

Durch die Thematisierung der gewünschten Veränderungen kann zusätzlich zur aktuellen ZB für Walderholung eine Obergrenze für die ZB für Walderholung abgeschätzt werden, da auch durch Verbesserungen im Wald keine höheren ZB erwartet werden können und die gefundenen Werte somit vermutlich eine Obergrenze darstellen. Gleichzeitig lässt sich

inhaltlich sagen, dass die genannten Änderungen zwar gewünscht sind, aber zu keinem höheren Erholungswert führen. Fragte man die Befragten gesondert nach einer ZB für bspw. Naturschutz, käme man vermutlich auf andere Werte, da Naturschutz selbst nicht unattraktiv ist, sondern nach den hier vorliegenden Ergebnissen nur keinen steigernden Einfluss auf die Erholung hat.

## **2.2 Zweiter Artikel „Preferences for forest structural attributes in Germany – evidence from a Choice Experiment“ (Weller und Elsasser, 2018, Forest Policy and Economics)**

Dieser Artikel wurde von Priska Weller (Federführung) und Dr. Peter Elsasser gemeinsam auf Englisch geschrieben und im April 2018 bei der peer-reviewten Zeitschrift „Forest Policy and Economics“ veröffentlicht. Priska Weller hat die Auswertungen durchgeführt, Dr. Peter Elsasser trug zu den Schlussfolgerungen bei. Das Befragungskonzept wurde von den Autoren gemeinsam entwickelt.

### **Zusammenfassung**

Um acht verschiedene Waldleistungen umweltökonomisch zu bewerten wurde 2013 eine Bevölkerungsbefragung mit einem DCE durchgeführt. 2932 zufällig aus dem online-Panel eines Befragungsinstituts ausgewählte Personen beantworteten den Fragebogen vollständig. Die Befragten waren auf zwei fast gleich große Stichproben verteilt (1465 Personen in Stichprobe 1, 1467 Personen in Stichprobe 2). Das zu bewertende Gut ist eine Landschaft, charakterisiert durch unterschiedliche Merkmale, die sich auf Eigenschaften des Waldes und teilweise auch der Landschaft beziehen. Die Merkmale sind Veränderungen am Anteil von Waldflächen an der Landschaft (a), am sichtbaren Feld-/Wald-Mosaik (b), Verbreitung von Unterwuchs im Wald (c), veränderte Anteile von Nadelbäumen (d) und nichtheimischen Bäumen (f), zeitliche Verschiebung der Holzernte (e) sowie Aufhebung der Nutzung einzelner Waldflächen (g), und Steigerungen der Biodiversität (h). In dem Auswahlexperiment werden auf neun nacheinander präsentierten Auswahlkarten fünf Landschaftsmerkmale plus ein Preismerkmal präsentiert, von denen zwei Landschaftsmerkmale und das Preismerkmal in beiden verwendeten Stichproben vorkommen. So können acht Landschaftsmerkmale untersucht werden, ohne den Befragten zu viele Merkmale gleichzeitig zu präsentieren (Merkmale in Stichprobe 1: a, b; c, d, e, Preis; Merkmale in Stichprobe 2: a, b; f, g, h, Preis). Jedes Landschaftsmerkmal hat zwei Ausprägungen neben dem Status Quo, von dem ausgehend die Veränderungen gegeneinander

abgewogen werden sollen. Dieser aktuelle Zustand charakterisiert auch die Ausweichmöglichkeit: eine Landschaft mit ausschließlich „Status Quo“-Level, die zu Kosten von 0€ gewählt werden kann. Das Preismerkmal hat sechs Ausprägungen neben dem Status Quo.

Es werden zwei unterschiedliche Modelle zur Errechnung der Präferenzen verwendet. Die Präferenzen werden im Multinomialen Logit-Modell (MNL; unter der Annahme einheitlicher Präferenzen der Befragten) und im Mixed Logit-Modell (unter der Annahme von Präferenzheterogenität der Befragten; Verteilungsannahmen der Parameter beziehen sich auf die ZB (im „Willingness to pay space“, WTP space)) berechnet. Insgesamt sind die Ergebnisse in den beiden Ansätzen ähnlich, insbesondere bei der Betrachtung der Vorzeichen, also ob positive oder negative Präferenzen vorliegen. Dabei wird in den Modellergebnissen des Mixed Logit-Modells aber auch sichtbar, dass bei vielen der Merkmale die Ergebnisse durch eine hohe Heterogenität gekennzeichnet sind und dass dies auch ein Treiber der Modellqualität ist. Resultierend daraus wird die ZB aus den Koeffizienten des Mixed Logit-Modells im WTP space berechnet.

Die monetären Bewertungen der Befragten beziehen sich auf stufenweise Änderungen vom Status Quo hin zu jeweils einer der anderen Merkmalsausprägungen (dummy) und zeigen, dass nur wenige der angebotenen Änderungen weg vom Status Quo der Landschaft positive jährliche ZB generieren. Dies gilt für eine Steigerung des Anteils von Waldflächen (a), jede Steigerung der Biodiversität (h), eine zeitliche Verschiebung der Holzernte um 20 Jahre (e) und eine Aufhebung der Nutzung von Wald auf 10% der Flächen (g). Die verbleibenden Änderungen weg vom Status Quo der Landschaft sind mit unterschiedlich hohen negativen ZB belegt. Dies gilt für eine Reduktion des Anteils von Waldflächen (a), Vergrößerungen und Verkleinerungen des Feld-/Wald-Mosaiks (b), die Reduktion von Unterwuchs (c), auf 70% veränderte Anteile von Nadelbäumen (d) und Erhöhungen wie Reduktionen von nichtheimischen Bäumen (f), sowie eine Abschaffung von ungenutzten Waldflächen (g). Andere Veränderungen haben nicht zu statistisch signifikanten ZB geführt: eine Veränderung des Nadelbaumanteils auf 30% (d), eine Vergrößerung der Unterwuchsmenge (c) und eine größere zeitliche Verschiebung der Holzernte um 30 Jahre (e). Die Erhöhung des Anteils von Bäumen aus anderen Ländern ist unter den höchsten negativen Werten (-121€), ebenso wie die negative ZB für eine hypothetische Reduktion des Waldanteils (-122€ (Stichprobe 1) und -110€ (Stichprobe 2)). Besonders hohe positive ZB werden für eine Erhöhung des Waldanteils

(44€ (Stichprobe 1) und 46€ (Stichprobe 2)) und eine Erhöhung der Biodiversität (mindestens 34€) errechnet. Dies entspricht den Erwartungen. Die absolute Höhe der ZB wird stark davon beeinflusst, dass die Modellberechnung im WTP space stattfindet: Das Modell im WTP space ergibt besonders niedrige ZB-Werte und eignet sich deshalb für konservative Bewertungen und deren Verwendung in politischen Kontexten.

### **Diskussion und Einordnung in den thematischen Rahmen**

Dieser Artikel stellt zum einen tatsächlich erhobene Werte für Zahlungsbereitschaften der Bevölkerung für einige Güter des Waldes dar, zum anderen zeigt er anhand von unterschiedlichen Modellspezifikationen die mögliche Bandbreite an Ergebnissen.

Die erhobenen ZB-Werte gelten pro Person und Jahr und beschreiben den Durchschnitt der Bevölkerung. Zunächst werden in über 50% der Auswahlentscheidungen Präferenzen für den Status Quo zum Ausdruck gebracht. In den verbleibenden Auswahlentscheidungen werden Eingriffe in die Baumartenausstattung ausdrücklich negativ bewertet, allerdings sind die negativen Werte für die unbeliebten Änderungen (mehr nichtheimische Baumarten, Nadelbaumanteil auf 70%) deutlich höher. Die ZB für Änderungen im Feld-/Wald-Mosaik werden in beide Richtungen recht ähnlich bewertet – als Erklärung bietet sich an, dass die Befragten im Durchschnitt zufrieden sind (manche wollen aber größere, manche kleinere Unterteilungen). Wie erwartet werden Reduktionen des Waldanteils negativ und Erhöhungen positiv bewertet. Für zwei Landschaftsmerkmale liegen für jede Ausprägung zwei ZB vor, einer aus Stichprobe 1 und einer aus Stichprobe 2. Für beide Ausprägungen des Merkmals Waldanteil überlappen die Konfidenzintervalle der jeweiligen ZB und deuten auf gleich große ZB hin. Dies gilt auch für eine Reduktion des Feld-/Wald-Mosaiks, nur bei einer Erhöhung des Feld-/Wald-Mosaiks unterscheiden sich die ZB in den beiden Stichproben. Dies mag zum einen auf leichte Unterschiede in den Entscheidungskontexten zurückzuführen sein, die durch die unterschiedlichen Merkmale generiert wurden, die zusätzlich präsentiert wurden, obwohl alle Merkmale sich auf Wald bezogen. Zum anderen mag dies auf Effekte der Stichprobenziehung für beide Stichproben zurückzuführen sein. Ein unerwartetes Ergebnis ist, dass leichte Erhöhungen der Biodiversität höhere positive Bewertungen erhalten als stärkere Erhöhungen der Biodiversität (allerdings bei überlappenden Konfidenzintervallen, die eine fast gleiche Größe der Werte andeuten).

Die Bewertungsergebnisse können mit den ebenfalls vorliegenden qualitativen Ergebnissen der Befragung verglichen werden: im Anschluss an die Beantwortung der Auswahlkarten wurden die Befragten gebeten für jedes Merkmal auf einer fünfstufigen Skala anzugeben, wie

wichtig es für die Auswahlentscheidungen insgesamt war. Für den Waldanteil wurde passend zur hohen ZB eine hohe Wichtigkeit des Merkmals für die Entscheidung angegeben, ebenso bei dem positiv bewerteten Merkmal zur Biodiversität und die Beendigung der Nutzung ausgewählter Waldflächen. Auch das Merkmal zum Feld-/Wald-Mosaik wird als wichtig angegeben, allerdings werden Änderungen in beide Richtungen negativ bewertet. Für eine Veränderung des Nadelbaumanteils auf 70% und eine Erhöhung des Anteils von nichtheimischen Baumarten liegen die höchsten negativen ZB vor, allerdings wurde dies von der Mehrzahl der Befragten als weder wichtig noch unwichtig eingestuft. Dies gilt auch für die Menge des Unterwuchses und das Erntealter, allerdings wurden diese Merkmale auch nicht auffallend hoch bewertet.

Die beiden untersuchten Modelle unterscheiden sich hinsichtlich ihrer Modellqualität. Berücksichtigung von Präferenzheterogenität aufseiten der Befragten erhöht die Modellqualität und das MNL ist unterlegen. Die Heterogenitätsanzeiger für Präferenzheterogenität der Befragten (SD der Koeffizienten im Mixed Logit-Modell) zeigen sich überwiegend signifikant. Heterogenität in den Präferenzen der Befragten liegt demzufolge vor (dies ist bei einer Landschaftsbewertungsstudie, die regional unterschiedliche Ausgangssituationen vorfindet, durchaus erwartungsgemäß). Diese zu berücksichtigen, erhöht dann folgerichtig die Modellqualität. Die Vorzeichen der Präferenzparameter basierend auf Modellberechnungen mit den beiden unterschiedlichen Annahmen zeigen aber in den meisten Fällen in die gleiche Richtung.

In der Literatur wird für das Mixed Logit-Modell im WTP space berichtet, wenig extreme ZB-Ergebnisse zu generieren (Train und Weeks, 2005; Scarpa et al., 2008; Hole und Kolstad, 2012). Es eignet sich somit besonders für konservative Schätzungen im Bereich der Politikberatung. Die unterschiedlichen Werte der unterschiedlichen Berechnungsmethoden lassen sich durch die zusätzliche Berücksichtigung von Präferenzheterogenität der Befragten im Mixed Logit-Modell, sowie die unterschiedlichen Annahmen über die Verteilungen der Parameter (WTP space) erklären.

### **2.3 Dritter Artikel „Stated and inferred attribute non-attendance in a design of designs approach“ (Weller, Oehlmann, Mariel, Meyerhoff, 2014, JoCM)**

Der Artikel wurde von Priska Weller (Federführung), Malte Oehlmann, Prof. Petr Mariel und Dr. Jürgen Meyerhoff gemeinsam auf Englisch geschrieben und nach der Präsentation auf der International Choice Modelling Conference 2013 in Sydney durch Priska Weller beim Journal

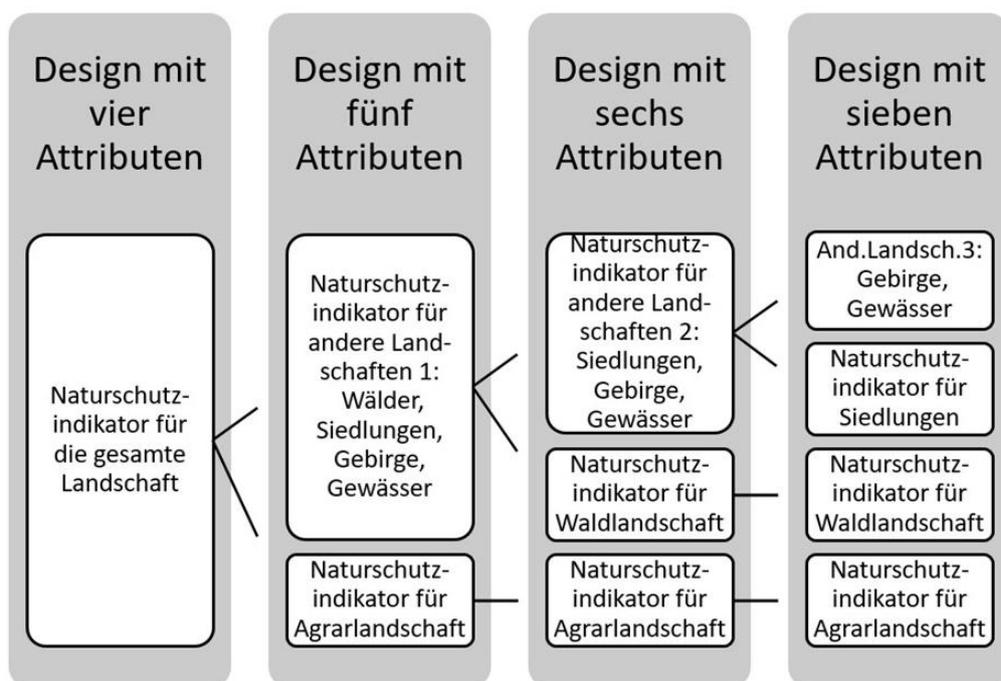
of Choice Modelling eingereicht und nach einem peer-review veröffentlicht. Das Befragungskonzept wurde von Priska Weller, Malte Oehlmann und Dr. Jürgen Meyerhoff entwickelt, die Berechnungen für die einzelnen Teile wurden hauptsächlich von Priska Weller, Prof. Petr Mariel und Dr. Jürgen Meyerhoff durchgeführt.

### **Zusammenfassung**

Bei der Durchführung von Choice Experimenten stellt sich die Frage, ob und wie die Befragten auf die Merkmale reagieren und ob sie alle angebotenen Merkmale in ihre Abwägungsentscheidung einbeziehen, oder Antwortheuristiken folgen, die die Entscheidung für eine Alternative erleichtern. Eine solche Antwortheuristik kann die Nicht-Berücksichtigung von einzelnen oder mehreren Merkmalen des Auswahl-experiments (englisch: attribute non-attendance, ANA) bei der Entscheidung für eine Alternative auf der Auswahlkarte sein. Hintergrund ist die Frage, inwiefern man (finanziell) aufwändige Bevölkerungsbefragungen ausnutzen kann, um möglichst viele Informationen zu erlangen, ohne dass es dabei zu merklichen Qualitätseinbußen durch kognitive Überforderung der Befragten kommt. Die bisher vorhandene Literatur hat die Berücksichtigung aller Merkmale als wichtig für die Befragungsergebnisse identifiziert (Hensher et al., 2005). Die Auslöser für die Befragten, einzelne Merkmale mehr oder weniger zu berücksichtigen wurden bisher nicht abschließend untersucht. Der vorliegende Artikel betrachtet die Variation in der Komplexität der Designerstellung nun systematisch als möglichen Auslöser für Nicht-Berücksichtigung von einzelnen Merkmalen durch die Befragten.

Es wurde eine weitere Bevölkerungsbefragung durchgeführt (diese ist auch die Grundlage für die Auswertung in Artikel vier). Das zu bewertende Gut war die Landschaft in der Wohnumgebung und die charakterisierenden Merkmale waren Waldanteil der Landschaft, Flächenumwandlung und unterschiedliche Indikatoren zum Naturschutz der Gegend. Der Naturschutzindikator muss für diese Befragung teilbar sein, um unterschiedliche Anzahlen von Merkmalen für die unterschiedlichen Stichproben zu erhalten. Dazu wurde der Ansatz des Vogelindikators (Achtziger et al., 2003; BMU, 2010) erweitert, so dass die vorhandenen Teilindikatoren für unterschiedliche Landschaftstypen additiv miteinander verknüpft und um eine Kategorie für „andere Landschaften“ erweitert wurden. Dieses Vorgehen ist inhaltlich angreifbar, da die Befragten (je nach Stichprobe) aufgefordert waren, zwischen zwei bis vier Arten von Naturschutzindikator abzuwägen (siehe Abbildung 2). Der Vorteil dieses Vorgehens liegt darin, dass kein zusätzliches Merkmal mit einem anderen Fokus (hier könnte

man sich beispielsweise ein Merkmal zum Anteil von Agrarflächen vorstellen) die Abwägung verzerrt. Nachteilig an der Verwendung eines aufgeteilten Naturschutzindikators ist, dass für die Befragten die Abwägung unterschiedlicher Naturschutzindikatoren gegeneinander wahrscheinlich nicht sinnvoll machbar ist. Da hier die inhaltliche Interpretation der Ergebnisse aber nicht im Fokus stand, sondern die methodische hinsichtlich der Berücksichtigung von Merkmalen, wurde dieses Vorgehen gewählt. Bereits in Artikel zwei wurde der Teilindikator für Wald-Habitate des Vogelindikators für die Bewertung der Biodiversität im Wald verwendet, allerdings nicht weiter unterteilt.



**Abbildung 2: Unterteilung des Naturschutzindikators** (übersetzt nach Meyerhoff et al., 2015)

Nach einem „Design of Designs“-Plan aus der Literatur (Hensher, 2004) wurde für diese Befragung das Design systematisch variiert, so dass 16 kleine Stichproben mit ca. 80 – 220 Befragten zur Verfügung stehen. Die Auswahlexperimente für die unterschiedlichen Stichproben variieren in unterschiedlichen Kombinationen der folgenden fünf Dimensionen: Auf 6 - 24 Auswahlkarten (in vier Schritten á sechs Karten) werden drei bis fünf Alternativen (inkl. Status Quo-Alternative) mit vier bis sieben Merkmalen (immer die Merkmale Waldanteil, Flächenumwandlung und Preismerkmal und zusätzlich eine Variante des Naturschutzindikators wie in Abbildung 2 dargestellt) präsentiert, die über zwei bis vier unterschiedliche Level und drei Levelspannweiten variieren (siehe Tabelle 1). Nach dem DCE

wurden die Befragten aufgefordert, auf einer fünfstufigen Skala (niemals, selten, manchmal, häufig, immer berücksichtigt; Scarpa et al. (2013) haben auch einen fünfstufigen Ansatz zur Erhebung von Nicht-Berücksichtigung von Merkmalen verwendet) ihre persönliche Berücksichtigung für jedes der einzelnen Merkmale anzugeben. Dies gibt die geäußerte Nicht-Berücksichtigung wieder. Die Stufen wurden in die Kategorien „Nicht-Berücksichtigung“ (ursprüngliche Antwortkategorien niemals, selten und manchmal) und „Berücksichtigung“ (ursprüngliche Antwortkategorien häufig und immer) für jedes Merkmal zusammengefasst. Mittels Chi-Quadrat-Test wurde dann das Verhältnis zwischen den binären Antwortkategorien und den unterschiedlichen Ausprägungen der fünf Designdimensionen statistisch untersucht.

Zusätzlich wurden zur rechnerischen Herleitung der Nicht-Berücksichtigung von einzelnen Merkmalen fünf mögliche Klassen definiert und die Wahrscheinlichkeiten für die jeweilige Klassenzugehörigkeit mittels latentem Klassenmodell (englisch: equality-constrained latent class model; ECLCM) für jeden Befragten in jeder der 16 Stichproben modelliert. In der ersten Klasse wurden alle jeweils präsentierten Merkmale berücksichtigt, in der zweiten bis vierten Klasse wurde jeweils eines der drei Merkmale Waldanteil, Flächenumwandlung und Preismerkmal nicht berücksichtigt und in der fünften Klasse wurde kein Merkmal berücksichtigt (dies entspricht einer zufälligen Auswahl, da die Entscheidung durch keines der Merkmale beeinflusst ist). Im weiteren Vorgehen in der Modellierung werden die Nutzegewichtungen der jeweils nicht berücksichtigten Merkmale als Null definiert, um die Nicht-Berücksichtigung im Modell einzubeziehen. Durch Gleichsetzung der Modellkoeffizienten der Merkmale über die Klassen hinweg wird erreicht, dass keine Präferenzheterogenität zwischen den Befragten abgebildet wird. Letzteres wird in Kauf genommen, da für die weitere Analyse die Wahrscheinlichkeiten für die Zugehörigkeit zu den fünf definierten Klassen erheblich sind. Diese sagen nämlich aus, in welchem Maße welche Merkmale berücksichtigt oder nicht berücksichtigt wurden (abhängig von der oben genannten Abgrenzung der einzelnen Klassen). Anschließend wurde der Einfluss der Designdimensionen auf die Wahrscheinlichkeiten mittels Spearmans Rangkorrelationskoeffizient statistisch untersucht. Für das Merkmal Naturschutzindikator werden weder die geäußerte Berücksichtigung, noch die Wahrscheinlichkeit einer Klassenzugehörigkeit betrachtet, da die vorliegenden Beobachtungszahlen für die einzelnen Unterteilungen durch die Designstruktur nicht einheitlich sind.

**Tabelle 1: „Design of Designs“** (Struktur der Tabelle nach Hensher, 2004, Ausprägungen der Anzahlen und Spannbreiten angepasst für Meyerhoff et al., 2015)

Design	Anzahl Auswahlkarten	Anzahl Alternativen	Anzahl Merkmale	Anzahl Levels	Spannbreite Levels
1	24	4	5	3	Basis
2	18	4	5	4	Basis +20%
3	24	3	6	2	Basis +20%
4	12	3	6	4	Basis
5	6	3	4	3	Basis +20%
6	24	3	4	4	Basis -20%
7	6	4	7	2	Basis -20%
8	12	5	4	4	Basis +20%
9	24	5	4	4	Basis
10	6	5	7	3	Basis +20%
11	6	4	6	4	Basis -20%
12	12	5	5	2	Basis -20%
13	18	4	7	2	Basis
14	18	3	4	3	Basis -20%
15	12	3	5	2	Basis
16	18	5	6	3	Basis -20%

In Bezug auf die insgesamt geäußerte Nicht-Berücksichtigung ist nur ein Zusammenhang mit der Designdimension „Anzahl von Merkmalen“ festzustellen. Die Berücksichtigung der Merkmale wird höher angegeben, wenn weniger Merkmale in einem Design enthalten waren. In Bezug auf die geäußerte Nicht-Berücksichtigung für einzelne Merkmale wurde speziell für das Merkmal Waldanteil im Durchschnitt häufige Berücksichtigung angegeben, für das Preismerkmal wurde im Durchschnitt manchmal berücksichtigt angegeben und der Wert für das Merkmal Flächenumwandlung lag dazwischen. Der Einfluss der Designdimensionen auf die geäußerten binär codierten Nicht-Berücksichtigungswerte für jedes Merkmal einzeln ist nur in fünf Fällen signifikant. Je höher die Anzahl der Auswahlkarten, desto geringer die angegebene Häufigkeit von Berücksichtigung für einzelne Merkmale (hier: Flächenumwandlung). Je höher die Anzahl von Alternativen, desto niedriger die angegebene Häufigkeit von Berücksichtigung (hier: Waldanteil und Flächenumwandlung). Auch eine höhere Anzahl von Levels führt zu einer geringeren Häufigkeit von Berücksichtigung (hier: des Preismerkmals) und eine größere Spannbreite der Levels führt zu einer geringeren Häufigkeit von Berücksichtigung vs. Nicht-Berücksichtigung (hier: Waldanteil).

In Bezug auf die hergeleitete Nicht-Berücksichtigung lässt sich aus dem latenten Klassenmodell ableiten, wie schon bei der angegebenen Nicht-Berücksichtigung, dass nur von einem geringen Anteil der Befragten alle Merkmale berücksichtigt wurden. Je nach Design betrug die Gruppengröße 17% bis 33%. Die Gruppengröße der Befragten, die kein Merkmal

berücksichtigen und zufällige Wahlen treffen, lag zwischen 5% und 24%. Der Einfluss der Designdimensionen ist hier hauptsächlich in zwei Dimensionen sichtbar. Zum einen wird bei einer höheren Anzahl von Auswahlkarten das Merkmal Waldanteil weniger berücksichtigt. Zum anderen hat eine zunehmende Anzahl der Alternativen Einfluss auf fast alle Klassenwahrscheinlichkeiten, allerdings in unterschiedliche Richtungen: die Wahrscheinlichkeit, alle Merkmale zu berücksichtigen, sowie die Wahrscheinlichkeit, den Waldanteil nicht zu berücksichtigen, nehmen ab. Die Wahrscheinlichkeiten, das Preismerkmal nicht zu berücksichtigen und kein Merkmal zu berücksichtigen, nehmen zu.

Nach der Analyse lässt sich festhalten, dass in beiden Ansätzen kein starker Zusammenhang zwischen den Designdimensionen und der Berücksichtigung von Merkmalen gefunden wurde. Die Ergebnisse der beiden Ansätze stimmen weitgehend überein in ihrer Einschätzung der Zusammenhänge zwischen geäußerter oder hergeleiteter Nicht-Berücksichtigung und den Dimensionen des Choice Experiments. Aus forschungstheoretischer Sicht ist es ein gutes Ergebnis, dass überwiegend von wenig Einfluss des Designs des DCE auf die Ergebnisse ausgegangen werden kann. Bei manchen Dimensionen gibt es allerdings einen sichtbaren Einfluss auf die Ergebnisse. Dies betrifft zum einen die Anzahl an angebotenen Alternativen: Je höher diese sind, desto mehr sinkt die Berücksichtigung aller Merkmale. Zum anderen betrifft dies die Anzahl an Auswahlkarten: Je mehr angeboten werden, desto stärker nimmt die Nicht-Berücksichtigung aller Merkmale zu.

### **Diskussion und Einordnung in den thematischen Rahmen**

Bei der Untersuchung von geäußerter und hergeleiteter Nicht-Berücksichtigung von Merkmalen in DCE spielt ein Merkmal eine besonders große Rolle: das Preis-Merkmal. Dieses wurde von den Befragten nur eingeschränkt berücksichtigt, obwohl es für die Wissenschaftler in einer ökonomischen Untersuchung besonders relevant ist. Dies war in der geäußerten sowie der hergeleiteten Nicht-Berücksichtigung sichtbar: Auf der fünfstufigen Skala wurde das Preis-Merkmal als „manchmal“ berücksichtigt angegeben, und die Wahrscheinlichkeit der „Preis nicht berücksichtigt“-Klasse lag zwischen 11 und 40% je nach Stichprobe. Dies weist darauf hin, dass ein wesentlicher Anteil der Befragten den Preis der Alternativen nicht in die Entscheidung einbezogen hat. Dies ist ein unerwartetes Ergebnis, allerdings finden sich in der Literatur Belege, die dieses Ergebnis bestätigen (Campbell et al. (2008) fanden, dass 69% der Befragten das Preismerkmal berücksichtigen; Kragt, 2013). Ein Teil der Begründung ist in der Höhe der angebotenen Preise zu finden: solange diese unterhalb einer bestimmten Grenze liegen, scheinen sie nicht wichtig für die Befragten zu

sein. Dies wurde in der Vergangenheit auch in Fokusgruppen zu diesem Thema geäußert. In Stichproben, denen eine große Anzahl von Merkmalen präsentiert wurde, könnte auch der Fokus auf wichtigere Merkmale eine Rolle gespielt haben. Bei der geäußerten Nicht-Berücksichtigung des Preis-Merkmals könnte auch soziale Erwünschtheit eine Rolle gespielt haben und somit die Angaben zu Ungunsten des Preis-Merkmals verzerrt haben, allerdings ist dieses Phänomen auch in der hergeleiteten Nicht-Berücksichtigung sichtbar und somit wird diese Kritik in Frage gestellt.

In der Betrachtung der geäußerten Nicht-Berücksichtigung wird die Belastbarkeit der Angaben der Antwortenden zu ihrer subjektiven Berücksichtigung hinterfragt, allerdings werden auch Modelle mit hergeleiteter Nicht-Berücksichtigung kritisch gesehen. Unsere Ergebnisse widersprechen sich hier nicht. Darüber hinaus wird die Belastbarkeit der ECLCM von manchen Studien kritisiert (Hess et al., 2013; Hensher et al., 2013), da diese keine Präferenzheterogenität aufseiten der Befragten berücksichtigen. Dadurch werden alle sichtbaren Schwankungen in den Ergebnissen der Nicht-Berücksichtigung der Merkmale zugeschlagen, was zu einer Überschätzung führen kann und tatsächlich unterschiedliche Präferenzen als Ursache haben könnte.

Insgesamt ist es ein interessantes Ergebnis, dass die Nicht-Berücksichtigung von Merkmalen nicht sehr stark von den Designdimensionen beeinflusst wird. Dies gibt Wissenschaftlern die Möglichkeit, mehr Dimensionen in die Befragung einzubeziehen, da die Vergrößerung des DCE überwiegend keinen Einfluss auf die Berücksichtigung der Merkmale hat. Bei zwei Dimensionen des DCE ist dies allerdings nicht der Fall: Zum einen ist die Anzahl der Alternativen ein Haupttreiber von Nicht-Berücksichtigung, ersichtlich aus der Verwendung zweier Ansätze zur Analyse von Nicht-Berücksichtigung von Merkmalen durch Befragte in DCE: geäußertes und hergeleitetes. Es zeigt sich unter anderem, dass mit mehr angebotenen Alternativen die Antworten „zufälliger“ werden, da die Gruppe der „kein Merkmal berücksichtigt/zufällige Auswahl“ mit mehr Alternativen größer wird. Dies ist vermutlich auf die mit mehr Anstrengung verbundene Abwägung zusätzlicher Alternativen zurückzuführen. Zum anderen führt die Dimension „Anzahl Auswahlkarten“ nach beiden Analysen zu mehr Nicht-Berücksichtigung von Merkmalen. Eine Erklärung könnte die größere Anstrengung der Befragten sein, die es bedeutet, mehr Auswahlkarten nacheinander zu beantworten. Entgegen den Erwartungen zeigt eine höhere Anzahl von Merkmalen keine Auswirkungen auf die

Nicht-Berücksichtigung von Merkmalen. Die Anzahl und Spannbreite von Levels zeigen keinen signifikanten Einfluss auf die Nicht-Berücksichtigung von Merkmalen.

Die ersten beiden genannten Dimensionen sollten also nach den vorliegenden Ergebnissen in zukünftigen DCE nicht zu stark ausgereizt werden. Insgesamt ist die Gefahr der Nicht-Berücksichtigung einzelner Merkmale aber nicht sehr groß und in einer Befragung sind sehr viele Informationen auf einmal von den Befragten erlangbar. Für diese beiden als einflussreich identifizierten Dimensionen „Anzahl Alternativen“ und „Anzahl Auswahlkarten“ können in zukünftigen Untersuchungen die Grenzwerte analysiert werden, an denen die Nicht-Berücksichtigung besonders ausgeprägt ist.

## **2.4 Vierter Artikel „The Influence of Design Dimensions on Stated Choices in an Environmental Context“ (Meyerhoff, Oehlmann, Weller, 2015, ERE)**

Der Artikel wurde von Dr. Jürgen Meyerhoff (Federführung), Malte Oehlmann und Priska Weller gemeinsam konzipiert und auf Englisch geschrieben und nach einem peer-review bei der Zeitschrift „Environmental and Resource Economics“ veröffentlicht. Die Berechnungen wurden von Dr. Jürgen Meyerhoff und Malte Oehlmann durchgeführt.

### **Zusammenfassung**

Derselbe Datensatz, auf dem auch die Auswertung zur Nicht-Berücksichtigung von Merkmalen (in Artikel drei) basiert, wurde auch für diesen Artikel verwendet. Er besteht aus 16 kleineren Unterstichgruppen, die nach einem „Design of Designs“-Plan nach Hensher (2004) erstellt wurden und in denen fünf Designdimensionen variiert werden (siehe Tabelle 1).

Die erste Forschungsfrage in der Auswertung bezieht sich auf die Abbruchrate der Befragten, in Abhängigkeit von der Ausgestaltung des jeweiligen Designs. Die Untersuchung bezieht sich auf das vorzeitige Beenden der Befragung durch eine Person, wobei die Abbrüche vor, während und nach der Beantwortung der Auswahlkarten stattfinden und während der Befragung fragegenau erfasst wurden.

Im Ergebnis lässt sich feststellen: Die Wahrscheinlichkeit, ob eine befragte Person die Befragung vorzeitig abbricht, hängt von ihren soziodemografischen Charakteristika ab (ältere Personen brechen eher ab als jüngere und Männer eher als Frauen), sowie einer steigenden Anzahl an Auswahlkarten, einer steigenden Anzahl an Merkmalen, die gegeneinander abgewogen werden, sowie der Anzahl von angebotenen Alternativen (bei fünf wird eher abgebrochen als bei drei). Die Abbruchrate beträgt insgesamt 20,6% und lässt sich aus dem

Datensatz separat für Abbrüche vor, innerhalb und nach dem DCE untersuchen. Dies zeigt, dass knapp die Hälfte der Abbrecher schon vor Beginn des DCE die Befragung abbricht (8,3%). Da vor dem DCE alle Fragebögen identisch waren, ist dies unabhängig von den Designdimensionen. Abbrüche, die vor dem DCE stattgefunden haben, sind noch nicht dem Design anzulasten und werden deshalb im Folgenden nicht berücksichtigt. Gut die Hälfte der Abbrecher bricht innerhalb des DCE ab (10,8%), und dort unterschiedlich über die (gruppierten) Auswahlkarten verteilt: 53,7% der Abbrecher brechen innerhalb der ersten 6 Auswahlkarten die Befragung ab. Zwischen den letzten Auswahlkarten (19-24) brechen im Vergleich dazu nur noch 8,7% der abbrechenden Befragten das Interview ab. Insgesamt fällt die Abbruchrate so von 6,3% der Befragten in Designs mit bis zu 6 Auswahlkarten auf 4,4% der Befragten in Designs mit 19-24 Auswahlkarten. Vorsicht bei der Interpretation der Ergebnisse: in der Gruppe derer, die innerhalb der ersten sechs Auswahlkarten abbrechen sind alle Befragten des DCE, während in der Gruppe derer die innerhalb der letzten 19-24 Auswahlkarten abbrechen, nur noch die Personen sein können, denen ein Design mit so vielen Auswahlkarten zugewiesen wurden (etwa ein Viertel der Befragten im DCE (454/1947 Personen)). Es lässt sich vermuten, dass manche Personen diese Art von Befragung ablehnen und dann abbrechen, die restlichen teilnehmenden Befragten wiederum lehnen Auswahlexperimente nicht prinzipiell ab und nehmen auch dann noch teil, wenn sie eine große Anzahl von Auswahlkarten zu absolvieren haben. Nach dem DCE brechen kaum noch Befragte das Interview ab (1,6%).

Die zweite Forschungsfrage untersucht mittels eines multinomialen Logit-Modells (MNL) und zweier heteroskedastischer Logit-Modelle (HL1 und HL2), inwiefern die Auswahlen, und insbesondere die Sicherheit der Auswahl, durch die Designdimensionen beeinflusst sind. Aus HL1 geht hervor, dass das Verhältnis zwischen der Varianz des Fehlerterms und der ersten Designdimension Anzahl Auswahlkarten konvex ist. Zunächst ist dies auf einen Lerneffekt zurückzuführen, der die Fehlervarianz reduziert, während später im Verlauf der Beantwortung der Auswahlkarten Ermüdung einsetzt und die Fehlervarianz wieder erhöht. Dies wird auch in der Literatur, z.B. von Caussade et al. (2005), so beschrieben. In Bezug auf die Dummies für die Designs mit unterschiedlichen Alternativen-Anzahlen zeigt sich, dass Designs mit vier Alternativen eine niedrigere Fehlervarianz aufweisen als Designs mit drei Alternativen, während der Dummy für fünf Alternativen keine signifikanten Ergebnisse erbringt. Auch hier decken sich unsere Ergebnisse wieder mit denen von Caussade et al. (2005). Die jeweilige Anzahl von Merkmalen und Merkmalsausprägungen beeinflussen beide die Fehlervarianz und

auch die Bandbreite der Merkmalsausprägungen hat einen signifikanten Einfluss: die Fehlervarianz sinkt in Designs mit schmalerer Bandbreite und steigt in Designs mit breiterer Bandbreite. Vergleiche der Alternativen sind leichter durchzuführen wenn die Bandbreiten der Merkmalsausprägungen schmaler sind. Auch hier decken sich die Ergebnisse wieder mit denen von Caussade et al. (2005). Von den soziodemografischen Erklärungsvariablen sind die getesteten Alter, Geschlecht und Bildung signifikant. Männliche Befragte zeigen eine größere Fehlervarianz beim Beantworten, während eine höhere Bildung eine geringere Fehlervarianz beim Beantworten nach sich zieht. HL2 zeigt anhand von Dummies, die die Position jeder Auswahlkarte anzeigen, die Veränderung der Fehlervarianz über die aufeinanderfolgenden Auswahlkarten. So sind auch Aufschlüsse über mögliche Lern- und Ermüdungseffekte im Verlauf der Beantwortung möglich. Aufgrund der Werte der Fehlervarianz über die verschiedenen Auswahlkarten hinweg lässt sich weder ein Lern- noch ein Ermüdungseffekt aus diesem Datensatz nachweisen.

Die aus den Modellen errechneten Koeffizienten deuten darauf hin, dass die Befragten im Durchschnitt einen größeren Waldanteil sowie höhere Werte beim Naturschutzindikator bevorzugen, während sie weniger Flächenumwandlung bevorzugen. Die aus dem MNL errechneten durchschnittlichen jährlichen ZB-Werte betragen 2,71€ für eine einprozentige Änderung des Waldanteils, -1,38€ für eine einprozentige Änderung der Flächenumwandlung und zwischen 0,44€ und 1,55€ für marginale Änderungen des Naturschutzindikators. Dabei erhält der Naturschutzindikator für die gesamte Landschaft die höchste ZB von 1,55€, höher als die Naturschutzindikatoren für die einzelnen Landschaftstypen. Die Ergebnisse für die auf niedrigerer Ebene aggregierten Naturschutzindikatoren zeigen Abstufungen für die unterschiedlichen Aggregationsniveaus. Dabei werden in den meisten Fällen auch die Naturschutzindikatoren mit mehr Landschaften höher bewertet als die mit weniger. Dieses Ergebnis entspricht den Erwartungen und unterstreicht die Plausibilität der Ergebnisse.

Es wurde untersucht, ob die Berücksichtigung des Einflusses der Designdimensionen auf die Fehlervarianz die Zahlungsbereitschaftswerte verändert. Dies ist nicht der Fall, da die Modelle sich nicht signifikant unterscheiden.

## **Diskussion und Einordnung in den thematischen Rahmen**

Das erste Ergebnis dieses Artikels ist, dass die Abbruchrate mit der Anzahl von Auswahlkarten und der Anzahl von Merkmalen steigt. Generell sind hohe Abbruchraten aber immer im Spannungsfeld der zusätzlich von den Befragten erlangbaren Antworten zu sehen

(Louviere et al., 2013). In einer persönlichen Befragung sind die Befragten unter Umständen auch zu motivieren bis zu 26 Auswahlkarten zu beantworten (Czajkowski et al., 2014). Die Abbruchrate hängt aber auch von einer generellen Vorliebe für Auswahlkarten ab. Wenn diese fehlt, wird die Befragung auch nach ganz wenigen Auswahlkarten abgebrochen.

Die Ergebnisse zur Frage des Einflusses der Designdimensionen auf die Fehlervarianz sind generell wie bei Caussade et al. (2005), aber drei Hauptunterschiede werden sichtbar: zum einen sind in dieser Befragung weniger Ermüdungseffekte sichtbar, außerdem ändert sich bei Betrachtung des Einflusses der einzelnen zusätzlichen Auswahlkarten die Entscheidungssicherheit nicht stark nach der ersten Auswahlkarte, und im vorliegenden Artikel wurden zwei signifikante Interaktionen gefunden. Die Position der Auswahlkarte und Anzahl von Alternativen hängen zusammen, so dass bei vier Alternativen die Fehlervarianz von einem anfänglich niedrigen Wert langsam steigt. Desweiteren hängen die Anzahl von Merkmalen und die Anzahl von Merkmalsausprägungen zusammen, wobei die ursprüngliche Fehlervarianz sinkt, wenn die Anzahl von Merkmalsausprägungen steigt.

Die Zahlungsbereitschaften in den untersuchten Modellen sind plausibel und unterscheiden sich nicht sehr stark, auch wenn im HL die Fehlervarianz berücksichtigt wird. Aufgrund zweier Punkte bedarf die Interpretation der Ergebnisse einer gewissen Vorsicht: Zum einen ist das als Zahlung in einen Fonds ausgestaltete Preismerkmal weniger anreizkompatibel als beispielsweise Zahlungen für private Güter, und zum anderen wird in diesen Modellen die Präferenzheterogenität aufseiten der Befragten nicht berücksichtigt. Ihr Vorliegen ist bei einer deutschlandweiten Befragung zu Landschaften nicht unwahrscheinlich und könnte die Ergebnisse stark verändern.

### **3. Diskussion und Schlussfolgerungen**

Die erste Forschungsfrage, die sich auf die kritische Anwendung der Methoden CVM und DCE und die Ableitung von Bewertungsergebnisse für einzelne Leistungen des Waldes bezieht, wird in den Artikeln eins und zwei behandelt. Dabei zeigt sich, dass die ZB für Walderholung mit ca. 27€ im niedrigeren, aber trotzdem eindeutig positiven, Bereich der ZB für einzelne Waldmerkmale aus Artikel zwei liegt. Die niedrigsten positiven Zahlungsbereitschaften aus Artikel zwei liegen bei ca. 11€ für eine Erhöhung der ungenutzten Waldflächen und ca. 25€ für eine Verschiebung des Erntezeitpunkts. Zahlungsbereitschaften für eine Erhöhung des Waldanteils und eine Erhöhung der Artenvielfalt liegen mit 44-46€ bzw. 34€ darüber.

Die zweite Forschungsfrage, die sich auf die Untersuchung methodischer Aspekte der DCE bezieht, wird in den Artikeln drei und vier behandelt. Die Kriterien, die dafür verwendet werden, sind der Einfluss von Designdimensionen auf Nicht-Berücksichtigung von Merkmalen, Abbruchquote und Fehlervarianz. Diese können nun in begrenztem Maße auch auf die Ergebnisse aus dem DCE in Artikel zwei angewandt werden und so ihre Belastbarkeit eingeschätzt werden. Dies gilt zunächst für die Nicht-Berücksichtigung einzelner Merkmale, in Abhängigkeit von den fünf untersuchten Designdimensionen. Es ist zu erwarten, dass die Nicht-Berücksichtigung nicht sehr ausgeprägt ist, da nur eine hohe Anzahl von Auswahlkarten Auswirkungen auf die Nicht-Berücksichtigung von Merkmalen hat. Die Zahl der Auswahlkarten lag bei neun Karten, und die getesteten Zahlen lagen bei bis zu 24. Desweiteren wurde eine hohe Anzahl von Alternativen als Treiber für die Nicht-Berücksichtigung von Merkmalen identifiziert, und auch hier lag die Anzahl der verwendeten Alternativen bei drei, also der niedrigsten Ausprägung in Artikel drei. Designs mit fünf Alternativen können auch eine höhere Abbruchquote erzeugen. Die Fehlervarianz wird zwar von den Designdimensionen beeinflusst, aber die angegebene ZB ändert sich nicht, auch wenn man die Einflüsse der Designdimensionen bei der Berechnung der ZB berücksichtigt, wie Artikel vier ergab. Für Entscheider bei Abwägungen politischer oder wirtschaftlicher Fragen ist die ZB für die bewerteten Güter auch die entscheidende Information.

Die dritte Forschungsfrage, die sich auf die Möglichkeit der Positionierung öffentlicher Güter in Verhandlungen bezieht, wird im Folgenden behandelt. Nachdem nun aus der ersten Forschungsfrage Werte für Erholung und Landschaftsaspekte des Waldes vorliegen und anhand der Kriterien aus der zweiten Forschungsfrage die Belastbarkeit der Ergebnisse aus

Artikel zwei besser eingeschätzt werden kann, liegen erste Möglichkeiten zur Lösung von Marktversagen vor. Die Möglichkeit zur Einbeziehung dieser Güter in Abwägungsentscheidungen über Maßnahmen im Wald, die diese Güter beeinflussen, ist nun gegeben und im Anschluss können Ressourcen zur Erreichung der Ziele aus den Maßnahmen entsprechend verteilt werden. Auch für Güter, die bisher nicht bewertet wurden, liegen nun Möglichkeiten vor, sie (in Bezug auf drei hier betrachtete Kriterien) als belastbar zu bewerten.

Aus den Artikeln eins und zwei liegen Ergebnisse vor, die sehr ähnlich strukturiert sind (beide Befragungen beziehen sich auf den Wald in der Nähe des Wohnortes, für Artikel eins war dies nicht näher konkretisiert, in Artikel zwei war dies als ein 15km-Radius um den Wohnort konkretisiert) und einen Vergleich anbieten. Dabei ist Vorsicht geboten, da es sich bei den beiden bewerteten Gütern um zwei unterschiedliche handelt: Artikel eins bewertet Walderholung, während Artikel zwei die Ansprüche an das Aussehen und andere Charakteristika des Waldes behandelt. Dabei können gleiche Merkmale ganz unterschiedlich bewertet werden. Der Wert der Walderholung muss nicht steigen, wenn das Waldaussehen verändert wird. Die Gemeinsamkeit der beiden Bewertungen bezieht sich darauf, dass nach einzelnen Waldmerkmalen gefragt wurde, allerdings einmal als Steigerungsmöglichkeit für den Erholungsnutzen und einmal als Unterteilung des Gutes „Naturschutz und Landschaft“. Aus Artikel eins ist bekannt, dass Befragte zwar Wünsche zur Veränderung des Waldes haben, deren Erfüllung den Erholungswert (die ZB für Walderholung) aber nicht steigert. Aus Artikel zwei ist andererseits bekannt, dass es signifikante Zahlungsbereitschaften für Veränderungen in der Waldausstattung gibt. Diese Diskrepanz lässt sich auf unterschiedliche Definitionen der bewerteten Güter und unterschiedliche Entscheidungskontexte in der Befragung zurückführen. In Artikel eins bezieht sich der Entscheidungskontext der Befragten auf die Erholung im Wald und wie der Erholungsnutzen sich steigern lässt. Die Möglichkeit zur Steigerung des Erholungsnutzens ist nicht groß, auch bei Durchführung von Maßnahmen zur Realisierung der Wünsche der Befragten, der Hauptfaktor ist zunächst die Verfügbarkeit von Wald überhaupt. In der Landschaftsbetrachtung in Artikel zwei werden dann Art und Aussehen des Waldes bewertet, und es zeigt sich, dass die Befragten Zahlungsbereitschaften für Veränderungen haben, auch wenn diese nicht mit dem Erholungsnutzen verknüpft sind. Den resultierenden Werten mag eine große nutzungsunabhängige Komponente zugrunde liegen, die an dieser Stelle die ZB stark beeinflusst.

Da die methodische Aussagefähigkeit der letzten beiden Artikel wichtiger war als die eher inhaltliche oder anwendungsbezogene Aussagefähigkeit, sind die inhaltlichen Interpretationen

mit mehr Vorsicht zu betrachten als die Ergebnisse aus Studien direkt zu anwendungsbezogenen Fragen. Dies zeigt sich in Artikel drei und vier beispielsweise an der Verwendung mehrerer Merkmale zum Naturschutzindikator, die zur Beantwortung methodischer Fragen nützlich waren und deshalb eingefügt wurden, umgekehrt für die Befragten aber inhaltlich schwer zu fassen waren.

Trotzdem sollen, zusätzlich zu den oben bereits diskutierten methodischen Ergebnissen, hier die inhaltlichen Ergebnisse zu Zahlungsbereitschaften aus Artikel vier nicht ausgelassen werden. Aus Artikel vier lassen sich die monetären Präferenzen für die Merkmale Waldanteil, Flächenumwandlung und Naturschutzindikator ablesen. Für eine Erhöhung des Waldanteils gibt es positive Präferenzen und einen positiven marginalen Zahlungsbereitschaftswert, andererseits eine negative marginale Zahlungsbereitschaft für Flächenumwandlung. Der Naturschutzindikator wird durchweg positiv bewertet, allerdings gibt es Abstufungen bei den unterschiedlichen Aggregationsniveaus für die unterschiedlichen Landschaftstypen. Insbesondere der Gesamtindikator wird höher bewertet (1,55€) als die anderen Indikatoren für einzelne Landschaftstypen. Bei den Indikatoren für einzelne Landschaftstypen wird der Indikator für Siedlungen (0,98€) leicht höher bewertet als die Indikatoren für Wald (0,95€) und Agrarflächen (0,88€). Bei den aggregierten Indikatoren ist der Wert des zweiten Indikators niedriger als der Wert des dritten, obwohl er mehr Landschaftstypen enthält (Siedlungen zusätzlich zu Bergen und Gewässer, siehe die obersten Felder der beiden rechten Säulen in Abbildung 2). Dies zeigt, dass die Befragten überwiegend ökonomisch sinnvoll antworten und Indikatoren mit mehr Landschaftstypen höher bewerten.

Die angegebenen ZB aus Artikel vier zu Waldanteil und Artenvielfalt lassen sich darüber hinaus mit den Ergebnissen aus Artikel zwei vergleichen. Trotz der unterschiedlichen Berechnungsarten (marginale Berechnung in Artikel vier und dummy-Codierung in Artikel zwei) deuten die Ergebnisse bei der Bewertung des Waldanteils darauf hin, dass grundsätzlich ein höherer Waldanteil positiv bewertet wird (2,71€ für marginale Erhöhung des Waldanteils; 44-46€ für eine schrittweise Erhöhung des Waldanteils) und die Zahlungsbereitschaften für Artenvielfalt/Naturschutzindikator sind in beiden Artikeln durchweg positiv (0,95€ für marginale Erhöhung des Wald-Indikators; 34€ für eine schrittweise Erhöhung des Wald-Indikators).

Da die Berücksichtigung der einzelnen Merkmale (bzw. ihre Einbeziehung in die Auswahlentscheidung) Aussagen über die Wichtigkeit der Merkmale erlaubt, können diese Informationen auch schon als Hinweis auf die Präferenzen und die zu erwartenden Zahlungsbereitschaften interpretiert werden. So lässt sich festhalten, dass in Artikel drei laut Angaben der Befragten das Merkmal zum Waldanteil insgesamt „häufig“ berücksichtigt wurde, was sich mit den Ergebnissen aus den Artikeln zwei und vier deckt, dass der Waldanteil an der Landschaft insgesamt ein wichtiges (hoch bewertetes) Merkmal für die Befragten ist. Das Merkmal zur Flächenumwandlung wurde etwas weniger berücksichtigt (zwischen „häufig“ und „manchmal“) und erhält eine negative marginale Zahlungsbereitschaft; diese ist in absoluten Zahlen etwas geringer als der Wert für den Waldanteil (-1,38€; 2,71€). Die überraschend seltene Berücksichtigung des Preismerkmals (zwischen „häufig“ und „manchmal“) wird auch in anderen Quellen erwähnt (Campbell et al., 2008, Preismerkmal wurde nur von 69% der Befragten berücksichtigt, weniger als jedes andere Merkmal in der Studie) und bestätigt. Dies führt aber nicht zu einer generellen Ablehnung des Konzepts der Zahlungsbereitschaften. Für die zahlreichen unterschiedlich unterteilten Merkmale zum Naturschutzindikator wurden keine mittleren Berücksichtigungsangaben berechnet, da sie unterschiedlichen Personenanzahlen vorgelegt wurden. Auch die in Artikel zwei qualitativ abgefragte Wichtigkeit der einzelnen Merkmale für die Auswahlentscheidung deutet darauf hin, dass oft als wichtig empfundene Merkmale mit einer höheren ZB bewertet werden.

Mit Vorliegen der Präferenzen der Bevölkerung und Werten für öffentliche Güter existiert nun die Möglichkeit auch nicht-marktliche Güter gegeneinander abzuwägen, dies muss nicht nur das Spannungsfeld mit privaten Gütern betreffen. Auch zwischen bspw. Erholungsleistungen und Biodiversitätsschutz können sich Zielkonflikte ergeben, die abgewogen werden müssen. So kann finanzielle Förderung zielgerichtet eingesetzt werden. Dies betrifft auch die Identifizierung und Umsetzung von politischen Strategien wie der Biodiversitätsstrategie oder der Nachhaltigkeitsstrategie, die zahlreiche Ziele verfolgen.

Für die Zukunft bleibt anzumerken, dass weitere öffentliche Güter bewertet werden müssen, um sie in politische Verhandlungen einbeziehen zu können. Dies betrifft auch Güter wie die Wasserreinigung oder die Klimaschutzleistung, für die bisher nur sehr uneinheitliche Werte vorliegen. Auswertungen auf tiefer regionalisierter Ebene tragen in Zukunft sicher auch zum Erkenntnisgewinn bei; dabei sollte auf die Unterscheidung zwischen den Präferenzen der

ländlich und städtisch lebenden Bevölkerung eingegangen werden, da mehr Menschen in Städten leben und entsprechend die Politik stärker beeinflussen. Die Waldpolitik beeinflusst andererseits aber die Menschen im ländlichen Raum stärker, so dass ein Ungleichgewicht zwischen Präferenzen und Lebenswirklichkeiten entsteht. Darüber hinaus können weitere methodische Untersuchungen zur Methode der DCE durchgeführt werden, die beispielsweise die Anreize für die Befragten, ihre wahren Präferenzen zu äußern, betreffen mögen. Letztendlich bleibt zu sehen, inwiefern politische Akteure Maßnahmen an den Präferenzen der Bevölkerung ausrichten wollen und für Entscheidungen darüber die vorliegenden Werte verwenden.

## Literaturverzeichnis

- Abildtrup, J., Garcia, S., Olsen, S. B., und Stenger, A., 2013. Spatial preference heterogeneity in forest recreation. *Ecological Economics*, 92, 67-77.
- Achtziger, R., Stickroth, H., und Zieschank, R., 2003. Nachhaltigkeitsindikator für den Naturschutzbereich-Informationspapier zur Artenauswahl: Kriterien und Vorgehensweise. Internetseite BfN, Bundesamt für Naturschutz: [http://www.bfn.de/fileadmin/MDB/documents/NHI\\_Naturschutzbereich\\_Infopapier\\_Manuskript.pdf](http://www.bfn.de/fileadmin/MDB/documents/NHI_Naturschutzbereich_Infopapier_Manuskript.pdf). Stand, 10, 2010.
- Adamowicz, W., Boxall, P., Williams, M., und Louviere, J., 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American journal of agricultural economics*, 80 (1), 64-75.
- Arentze, T., Borgers, A., Timmermans, H., und DelMistro, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review*, 39 (3), 229-244.
- Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., und Schuman, H., 1993. Report of the NOAA panel on contingent valuation. *Federal register*, 58 (10), 4601-4614.
- Balcombe, K., Burton, M., und Rigby, D., 2011. Skew and attribute non-attendance within the Bayesian mixed logit model. *Journal of Environmental Economics and Management*, 62 (3), 446-461.
- Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., et al., 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar.
- Bech, M., Kjaer, T., und Lauridsen, J., 2011. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health economics*, 20 (3), 273-286.

- Bergen, V., Löwenstein, W. und Olschewski, R., 2013. Forstökonomie - Volkswirtschaftliche Ansätze für eine vernünftige Umwelt- und Landnutzung. Verlag Franz Vahlen GmbH, München. 477 S.
- Bergen, V., Löwenstein, W., und Pfister, G., 1995. Studien zur monetären Bewertung von externen Effekten der Forst- und Holzwirtschaft. JD Sauerländer.
- Binder, S., Haight, R. G., Polasky, S., Warziniack, T., Mockrin, M. H., Deal, R. L., und Arthaud, G., 2017. Assessment and valuation of forest ecosystem services: State of the science review. Gen. Tech. Rep. NRS-170. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station., 47 S., 170, 1-47.
- BMELV, 2011. Waldstrategie 2020: Nachhaltige Waldbewirtschaftung - eine gesellschaftliche Chance und Herausforderung, Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz, Bonn. 36 S.
- BMEL, 2014. Der Wald in Deutschland. Ausgewählte Ergebnisse der dritten Bundeswaldinventur.
- BMU, 2007. Nationale Strategie zur biologischen Vielfalt. Reihe Umweltpolitik. 170 S.
- BMU, 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.
- BWaldG (Bundeswaldgesetz) vom 2. Mai 1975 (BGBl. I S. 1037), das zuletzt durch Artikel 1 des Gesetzes vom 17. Januar 2017 (BGBl. I S. 75) geändert worden ist. 13 S.
- Bösch, M., Elsasser, P., Franz, K., Lorenz, M., Moning, C., Olschewski, R., Rödl, A., Schneider, H., Schröppel, B., und Weller, P., 2018. Forest ecosystem services in rural areas of Germany: Insights from the national TEEB study. *Ecosystem Services*, 31, 77-83.
- Campbell, D., Hensher, D. A., und Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. *Journal of environmental planning and management*, 54 (8), 1061-1076.

- Campbell, D., Hutchinson, W. G., and Scarpa, R., 2008. Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. *Environment and Planning A*, 41 (1), 97-111.
- Carson, R. T., and Louviere, J. J., 2011. A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*, 49 (4), 539-559.
- Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., and Hensher, D. A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological*, 39 (7), 621-640.
- Chung, C., Boyer, T., and Han, S., 2011. How many choice sets and alternatives are optimal? Consistency in choice experiments. *Agribusiness*, 27 (1), 114-125.
- Costanza et al., 1997. The value of the world's ecosystem services and natural capital, *Nature* 387, 253-260.
- Czajkowski, M., Giergiczny, M., and Greene, W. H., 2014. Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Economics*, 90 (2), 324-351.
- Daily, G., ed., 1997. *Nature's services: societal dependence on natural ecosystems*. Washington, DC: Island Press. 392 S.
- Davis, R.K., 1963. *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*. Ph.D. Dissertation, Harvard University.
- Dellaert, B. G., Brazell, J. D., and Louviere, J. J., 1999. The effect of attribute variation on consumer choice consistency. *Marketing Letters*, 10 (2), 139-147.
- DeShazo, J. R., and Fermo, G., 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and management*, 44 (1), 123-143.

Destatis, 2016. Indikatorenbericht zur Nachhaltigkeitsstrategie.

Desvousges, W., Mathews, K., und Train, K., 2016. From Curious to Pragmatically Curious: Comment on “From Hopeless to Curious? Thoughts on Hausman's ‘Dubious to Hopeless’ Critique of Contingent Valuation”. *Applied Economic Perspectives and Policy*, 38 (1), 174-182.

Elsasser, P., 1996. Der Erholungswert des Waldes: monetäre Bewertung der Erholungsleistung ausgewählter Wälder in Deutschland. *Sauerländer*.

Elsasser, P., Hermann, E., und Hamilton, J., 2010. Landscape benefits of a forest conversion programme in North East Germany: results of a choice experiment. *Annals of Forest Research*, 53 (1), 3.

Elsasser, P., und Klein, C., 1994. Strategisches Verhalten als mögliche Fehlerquelle der Contingent Valuation Method (CVM). Zur Wertschätzung der Infrastrukturleistungen des Pfälzerwaldes. *Trippstadt, Mitt. aus der FVA Rheinland-Pfalz*, 27 (94), 111-128.

Elsasser, P., Meyerhoff, J., und Weller, P., 2016. An updated bibliography and database on forest ecosystem service valuation studies in Austria, Germany and Switzerland, Thünen Working Paper (No. 65).

Elsasser, P., und Weller, P., 2013. Aktuelle und potenzielle Erholungsleistung der Wälder in Deutschland. Monetärer Nutzen der Erholung aus Sicht der Bevölkerung. *Allgemeine Forst- und Jagdzeitung*, 184, 84-96.

Endres, A., 2007. *Umweltökonomie: Lehrbuch*. W. Kohlhammer Verlag.

Filyushkina, A., Agimass, F., Lundhede, T., Strange, N., und Jacobsen, J. B., 2017. Preferences for variation in forest characteristics: Does diversity between stands matter?. *Ecological Economics*, 140, 22-29.

- Galesic, M., 2006. Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of official statistics*, 22 (2), 313.
- Giergiczny, M., Czajkowski, M., Żylicz, T., und Angelstam, P., 2015. Choice experiment assessment of public preferences for forest structural attributes. *Ecological Economics*, 119, 8-23.
- Haab, T. C., Interis, M. G., Petrolia, D. R., und Whitehead, J. C., 2013. From hopeless to curious? Thoughts on Hausman's "dubious to hopeless" critique of contingent valuation. *Applied Economic Perspectives and Policy*, 35 (4), 593-612.
- Hanley, N., Mourato, S., und Wright, R. E., 2001. Choice modelling approaches: a superior alternative for environmental valuation?. *Journal of economic surveys*, 15 (3), 435-462.
- Hausman, J., 2012. Contingent valuation: from dubious to hopeless. *Journal of Economic Perspectives*, 26 (4), 43-56.
- Hensher, D. A., 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. *Journal of Transport Economics and Policy*, 38 (3), 425-446.
- Hensher, D. A., 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of applied econometrics*, 21 (6), 861-878.
- Hensher, D. A., Collins, A. T., und Greene, W. H., 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. *Transportation*, 40 (5), 1003-1020.
- Hensher, D. A., und Rose, J. M., 2009. Simplifying choice through attribute preservation or non-attendance: implications for willingness to pay. *Transportation Research Part E: Logistics and Transportation Review*, 45 (4), 583-590.

- Hensher, D. A., Rose, J., und Bertoia, T., 2007. The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies. *Transportation Research Part E: Logistics and Transportation Review*, 43 (2), 73-89.
- Hensher, D. A., Rose, J., und Greene, W. H., 2005. The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*, 32 (3), 203-222.
- Hensher, D. A., Rose, J. M., und Greene, W. H., 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation*, 39 (2), 235-245.
- Hensher, D. A., Stopher, P. R., und Louviere, J. J., 2001. An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application. *Journal of Air Transport Management*, 7 (6), 373-379.
- Hess, S., und Hensher, D. A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological*, 44 (6), 781-790.
- Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., und Caussade, S., 2013. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation*, 40 (3), 583-607.
- Hole, A. R., und Kolstad, J. R., 2012. Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empirical Economics*, 42 (2), 445-469.
- Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., et al., 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, 4 (2), 319-405.
- Kahneman, D., und Knetsch, J. L., 1992. Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management*, 22 (1), 57-70.
- Kling, C. L., Phaneuf, D. J., und Zhao, J., 2012. From Exxon to BP: Has some number become better than no number?. *Journal of Economic Perspectives*, 26 (4), 3-26.

- Kragt, M. E., 2013. Stated and inferred attribute attendance models: a comparison with environmental choice experiments. *Journal of Agricultural Economics*, 64 (3), 719-736.
- Krutilla, J. V., 1967. Conservation reconsidered. *The American Economic Review*, 57 (4), 777-786.
- Lancaster, K. J., 1966. A new approach to consumer theory. *Journal of Political Economy*, 74 (2), 132-157.
- Lindhjem, H., und Navrud, S., 2011. Using internet in stated preference surveys: a review and comparison of survey modes. *International Review of Environmental and Resource Economics*, 5 (4), 309-351.
- Löwenstein, W., 1994. Die Reisekostenmethode und die Bedingte Bewertungsmethode als Instrumente zur monetären Bewertung der Erholungsfunktion des Waldes: Ein ökonomischer und ökonometrischer Vergleich. Sauerländer, Frankfurt am Main.
- Louviere, J. J., Carson, R. T., Burgess, L., Street, D., und Marley, A. A., 2013. Sequential preference questions factors influencing completion rates and response times using an online panel. *Journal of Choice Modelling*, 8, 19-31.
- Louviere, J. J., Hensher, D. A., Swait, J. D., und Adamowicz, W., 2006. Stated choice methods: Analysis and applications. Cambridge University Press.
- Louviere, J. J., und Woodworth, G., 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research*, 350-367.
- Manski, C. F., 1977. The structure of random utility models. *Theory and decision*, 8 (3), 229-254.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, S. 105–142. Academic Press.

- Meyerhoff, J., Angeli, D., und Hartje, V., 2012. Valuing the benefits of implementing a national strategy on biological diversity—The case of Germany. *Environmental Science & Policy*, 23, 109-119.
- Meyerhoff, J., Oehlmann, M., und Weller, P., 2015. The influence of design dimensions on stated choices in an environmental context. *Environmental and resource economics*, 61 (3), 385-407.
- Mitchell, R. C., und Carson, R. T., 2005. Using surveys to value public goods: the contingent valuation method. Rff Press.
- MEA (Millennium Ecosystem Assessment), 2005. *Ecosystems and human well-being: synthesis*. Washington, DC: Island press.
- Mogas, J., Riera, P., und Bennett, J., 2006. A comparison of contingent valuation and choice modelling with second-order interactions. *Journal of Forest Economics*, 12 (1), 5-30.
- Möhring, B. und Mestemacher, U., 2009. Gesellschaftliche Leistungen der Wälder und der Forstwirtschaft und ihre Honorierung. *Landbauforschung - vTI Agriculture and Forestry Research Sonderheft 327*, 65–73.
- Olschewski, R., Bebi, P., Teich, M., Hayek, U. W., und Grêt-Regamey, A., 2012. Avalanche protection by forests—A choice experiment in the Swiss Alps. *Forest policy and Economics*, 17, 19-24.
- Pearce, D.W., 2001. The economic value of forest ecosystems. *Ecosyst. Health* 7, 284–296.
- Pindyck, R. S., und Rubinfeld, D. L., 2005. *Mikroökonomie* (6. Auflg.), Pearson, 960 S.
- Rolfe, J., und Bennett, J., 2009. The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics*, 68 (4), 1140-1148.

- Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82 (1), 34–55.
- Ryffel, A. N., Rid, W., und Grêt-Regamey, A., 2014. Land use trade-offs for flood protection: A choice experiment with visualizations. *Ecosystem services*, 10, 111-123.
- Sagebiel, J., Glenk, K., und Meyerhoff, J., 2017. Spatially explicit demand for afforestation. *Forest Policy and Economics*, 78, 190-199.
- Samuelson, P. A., 1954. The pure theory of public expenditure. *The review of economics and statistics*, 387-389.
- Scarpa, R., Gilbride, T. J., Campbell, D., und Hensher, D. A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European review of agricultural economics*, 36 (2), 151-174.
- Scarpa, R., Thiene, M., und Train, K., 2008. Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, 90 (4), 994-1010.
- Scarpa, R., Zanolini, R., Bruschi, V., und Naspetti, S., 2013. Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, 95 (1), 165-180.
- Swait, J., und Adamowicz, W., 2001. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, 86 (2), 141-167.
- Thurstone, L. L., 1927. A law of comparative judgment. *Psychology Review*, 34, 273–286.
- Train, K., 2009. *Discrete choice methods with simulation*. Cambridge University Press.

- Train, K., und Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In Applications of simulation methods in environmental and resource economics, S. 1-16. Springer, Dordrecht.
- Varian, H. R., und Buchegger, R., 2004. Grundzüge der Mikroökonomik (Vol. 6). München, Wien: Oldenbourg.
- Weisbrod, B. A., 1964. Collective-consumption services of individual-consumption goods. The Quarterly Journal of Economics, 78 (3), 471-477.
- Weller, P., und Elsasser, P., 2018. Preferences for forest structural attributes in Germany – Evidence from a choice experiment. Forest Policy and Economics, 93, 1-9.
- Weller, P., Oehlmann, M., Mariel, P., und Meyerhoff, J., 2014. Stated and inferred attribute non-attendance in a design of designs approach. Journal of choice modelling, 11, 43-56.

## **Anhang: wissenschaftliche Artikel**

1. Elsasser, P., und Weller, P., 2013. Aktuelle und potenzielle Erholungsleistung der Wälder in Deutschland. Monetärer Nutzen der Erholung aus Sicht der Bevölkerung. Allgemeine Forst- und Jagdzeitung, 184, 84-96.

# Aktuelle und potentielle Erholungsleistung der Wälder in Deutschland: Monetärer Nutzen der Erholung im Wald aus Sicht der Bevölkerung

Aus dem Johann Heinrich von Thünen-Institut, Institut für Forstökonomie<sup>1)</sup>

(Mit 3 Abbildungen und 6 Tabellen)

PETER ELSASSER<sup>\*)</sup> und PRISKA WELLER<sup>\*)</sup>

(Angenommen Oktober 2012)

## SCHLAGWÖRTER – KEY WORDS

*Ökosystemdienstleistungen; Erholung; monetäre Bewertung; Contingent Valuation; Deutschland; Wald.*

*Ecosystem services; recreation; monetary valuation; contingent valuation; Germany; forest.*

## 1. EINLEITUNG

Seit etwa 20 Jahren werden auch in Deutschland umweltökonomische Methoden genutzt, um den monetären Wert nicht marktgängiger Umweltleistungen der Wälder zu ermitteln. Mitte der 1990er Jahre erschien eine erste Serie umfangreicher Studien im Druck, welche den Erholungswert von Wäldern für Alltagsbesucher sowie für Urlauber zum Thema hatte; gleichzeitig wurden an diesem Beispiel auch etliche Details der verwendeten Bewertungsmethodik überprüft und weiterentwickelt (KLEIN und ELSASSER, 1994; LÖWENSTEIN, 1994; BERGEN *et al.*, 1995; LUTTMANN und SCHRÖDER, 1995; ELSASSER, 1996). Seither hat sich das Forschungsinteresse jedoch vermehrt anderen Umweltleistungen zugewendet, insbesondere dem monetären Nutzen von Naturschutzleistungen der Wälder für die Bevölkerung (s. ELSASSER *et al.*, 2009 für eine Übersicht der vorliegenden Studien). Der Wissensstand über Erholungswerte ist zwischenzeitlich recht aktualisierungsbedürftig geworden. Zudem informieren die bisherigen Studien primär über Erholungswerte im jeweils gegebenen Waldzustand; noch weitgehend unbekannt ist hingegen, ob und wie stark gezielte Veränderungen des Waldzustands den Erholungsnutzen der Bevölkerung beeinflussen könnten. Solche Veränderungen könnten sich beispielsweise durch vermehrte Berücksichtigung von Naturschutzbelangen in der Waldwirtschaft ergeben.

Ende 2011 wurde daher eine Bewertungsstudie durchgeführt, welche an die genannten Vorläuferstudien anknüpft und sowohl dem derzeit gegebenen Erholungswert der Wälder für Alltagsbesucher als auch dessen Verbesserungspotentialen nachgeht. Im Interesse der Vergleichbarkeit wurde – wie in den Vorläuferstudien – die Contingent Valuation Method (CVM) zur monetären Bewertung eingesetzt, und die zugehörige Konstruktion des hypothetischen Marktes für die aktuelle Erholungs-

leistung unverändert aus dem damals für Alltagsbesucher entwickelten Fragebogen übernommen (s. ELSASSER, 1996). Diese Konstruktion liegt auch verschiedenen weiteren Bewertungen der Walderholung zugrunde (KLEIN, 1994; BERNATH, 2006; KLEIBER, 2006). Insbesondere sollte durch die aktuelle Studie geprüft werden, welche Zahlungsbereitschaft heute für die Erholungsleistung der Wälder in Deutschland existiert, worauf Unterschiede der Zahlungsbereitschaften zurückgeführt werden können, sowie welche Verbesserungen die Bevölkerung in Hinblick auf die Erholungsleistung der Wälder wünscht und welche zusätzliche Zahlungsbereitschaft damit einhergehen könnte. Aus forschungsökonomischen Gründen interessierte zudem die Frage, ob sich die jeweiligen Zahlungsbereitschaften statt über persönliche Befragungen auch über online-Interviews eines Internet-Panels ermitteln lassen, ohne dass die Ergebnisse dadurch wesentlich beeinträchtigt würden.

## 2. BEWERTUNGSMETHODE

### 2.1. Contingent Valuation

Die hier verwendete Contingent Valuation Method zieht offenbarte Präferenzen zur monetären Bewertung heran; mit Hilfe von Interviews wird die maximale individuelle Zahlungsbereitschaft der Befragten für das zu bewertende Gut erhoben. Das Grundprinzip der Methode besteht darin, eine hypothetische Marktsituation für dieses Gut zu beschreiben und die Befragten anschließend nach ihrer maximalen Zahlungsbereitschaft für eine definierte Menge des Gutes unter den geschilderten Umständen zu fragen. Aus der Verteilung der Antworten können Mittelwert (sowie weitere Verteilungsmaße) der Zahlungsbereitschaft sowie auch die Nachfragekurve nach dem Gut geschätzt und anschließend gegebenenfalls auf die Bevölkerung hochgerechnet werden. Als Bewertungsmaß ergibt sich eine aggregierte Konsumentenrente (d.i. die Fläche unter der Nachfragekurve)<sup>2)</sup>, hier für das Gut „Erholung im Wald“ (s. z.B. BERGEN *et al.*, 2002:153 ff.).

Derartige Befragungen können im Detail auf unterschiedliche Weise gestaltet werden, was sich u.U. auch auf die Ergebnisse auswirkt (vgl. MITCHELL und CARSON, 1990; SÖDERQVIST und SOUTUKORVA, 2009). Im Interesse der Vergleichbarkeit war bereits für die Vorgänger-

<sup>1)</sup> Leuschnerstraße 91, D-21031 Hamburg.

<sup>\*)</sup> Korrespondierende Autoren:  
PETER ELSASSER. E-Mail: [peter.elsasser@vti.bund.de](mailto:peter.elsasser@vti.bund.de)  
PRISKA WELLER. E-Mail: [priska.weller@vti.bund.de](mailto:priska.weller@vti.bund.de)

<sup>2)</sup> Zur Vereinfachung wird hier nicht auf die Unterschiede zwischen MARSHALLSchen und HICKSSchen Bewertungsmaßen eingegangen.

studien ein im Kern einheitlicher Fragebogen für die Bewertung der Erholung im Wald entwickelt und zwischen forstökonomischen Instituten in der Bundesrepublik abgestimmt worden (ausführlicher dazu ELSASSER, 1996). Dieser wurde auch für die aktuelle Studie verwendet.

## 2.2. Fragebogaufbau

Im Fragebogen wird die Erholungsleistung des Waldes operationalisiert als das Recht, den bestehenden Wald zum Zwecke der Erholung zu betreten (entsprechend §14 BWALDG). Vor der eigentlichen Bewertung werden einige Einstimmungsfragen gestellt. Im vorliegenden Fall betrafen diese die Anzahl der Waldbesuche der vergangenen zwölf Monate in der Nähe des Wohnortes (11 Antwortvorgaben), die Entfernung zum zuletzt besuchten Wald (als offene Frage) sowie die beim letzten Waldbesuch hauptsächlich ausgeübte Tätigkeit (10 Antwortvorgaben einschließlich „Sonstiges“).

Daraufhin wurden die Befragten wie bereits in den Vorgängerstudien daran erinnert, dass das Erholungsangebot des Waldes kostenträchtig ist und teilweise staatlich, also durch „alle“ mitfinanziert wird. Anschließend wurde die maximale individuelle Zahlungsbereitschaft für eine fiktive Eintrittskarte erfragt, welche ein Jahr lang zum Besuch aller Wälder in der Nähe des Wohnortes berechtigen würde. Zur Absicherung der Antworten wurde anschließend nachgefragt, ob die Interviewpartner bei höheren Eintrittspreisen tatsächlich eher auf Waldbesuche verzichten würden als mehr zu bezahlen; daraufhin hatten sie Gelegenheit zur Revision des ursprünglich genannten Betrages. Bei einer Zahlungsbereitschaft von Null wurden die Gründe hierfür abgefragt, um echte Null-Zahlungsbereitschaften von Protestvoten<sup>3)</sup> unterscheiden zu können. Eine Bezahlkarte half bei der Bewertung (Abbildung 1). Sie zeigt 27 unterschiedliche Bewertungsvorschläge, die graphisch so arrangiert sind, dass etwaige Zufallsantworten einzelner Befragter ohne Einfluss auf die mittlere Zahlungsbereitschaft der Gesamtstichprobe bleiben (vgl. ELSASSER, 1996:65ff.). Aus diversen Begleitexperimenten der Vorgängerstudien lässt sich schließen, dass dieser Bewertungsansatz zu vergleichsweise „vorsichtigen“ Bewertungen der Erholungsleistung führt (ELSASSER, 1996, 1999); diese Bewertungen waren zudem in allen geprüften Fällen mit den offenbarten Präferenzen der Befragten kompatibel (d.h. sie hielten einer Kontrolle durch die methodisch unabhängige Reisekostenmethode stand. Ausführlicher dazu insbesondere LÖWENSTEIN, 1994; ELSASSER, 1996).

Im Anschluss an diese Bewertung der Walderholung im Status Quo wurden in der aktuellen Studie insgesamt fünf Zusatzfragen über mögliche Verbesserungen an den wohnortnahen Wäldern gestellt. Zunächst wurden die drei wichtigsten Verbesserungswünsche des jeweiligen Befragten ermittelt, gereiht nach Dringlich-

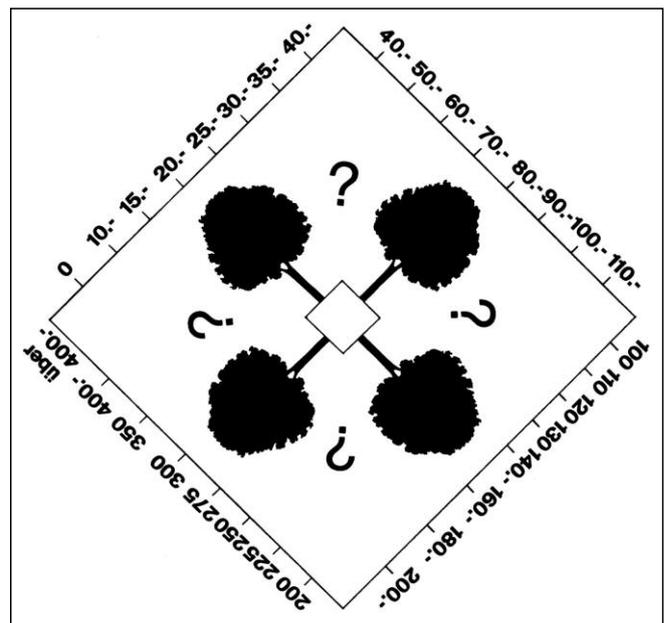


Abb. 1

Zur Bewertung eingesetzte Bezahlkarte.  
Payment card used for the valuation experiment.

keit (9 Antwortvorgaben einschließlich „Sonstiges“<sup>4)</sup>). Darauf wurde der Interviewpartner gefragt, ob er nach einer Einführung seiner drei wichtigsten Verbesserungswünsche mehr Waldbesuche unternehmen würde oder nicht. Die Fragen zur monetären Bewertung lauteten schließlich, ob sich die jährliche Zahlungsbereitschaft nach Einführung der wichtigsten Verbesserung erhöhen würde oder nicht, und wenn ja, auf wie viel; sowie wie hoch diese Zahlungsbereitschaft schließlich bei Umsetzung aller drei Verbesserungswünsche wäre.

Weitere soziodemographische Charakteristika wurden von dem beauftragten Befragungsunternehmen zur Verfügung gestellt (Geschlecht, Alter, Haushaltsgröße, Größe und Postleitzahl des Wohnorts, Angaben zu Schulbildung und Berufstätigkeit, monatliches Haushalts-Nettoeinkommen).

## 2.3. Stichprobe und Datenerhebung

Die Feldarbeit der aktuellen Studie (Stichprobenziehung und Interviews) wurde an ein spezialisiertes Unternehmen vergeben. Mit dem Ziel einer bevölkerungsrepräsentativen Abbildung wurde seitens dieses Unternehmens eine mehrstufige systematisch-randomisierte Stichprobe aus der Wohnbevölkerung Deutschlands über 14 Jahren gezogen. Insgesamt wurden so 2.535 Zielpersonen identifiziert (= Bruttostichprobe ohne stichprobenneutrale Ausfälle), von denen 1.011 bereit waren, sich befragen zu lassen (= Nettostichprobe; die

<sup>3)</sup> Als „Protestvoten“ werden geäußerte Zahlungsbereitschaften von Null bezeichnet, mit denen ein Befragter jedoch nicht seine wahre Zahlungsbereitschaft, sondern Protest gegen die Fragestellung kundtut.

<sup>4)</sup> ‚Verbesserte Sicherheit‘; ‚weniger Störungen durch andere Waldbesucher‘; ‚weniger Müll/aufgeräumterer Wald‘; ‚bessere/mehr Wege‘; ‚bessere/mehr Erholungsangebote (Wege, Parkplätze, Waldlehrpfade, Grillplätze usw.)‘; ‚größere Naturnähe‘; ‚mehr Laub- und Mischwälder anstelle von Nadelwäldern‘; ‚vielfältiger strukturierte Waldbestände anstelle von gleichförmigen‘; ‚Sonstiges, nämlich...‘

Stichprobenausschöpfung beträgt also 39,9%). Diese wurden zwischen dem 25. November und dem 6. Dezember 2011 in ihren Haushalten persönlich befragt (CAPI-Verfahren = Computer Assisted Personal Interviewing). Ergänzend dazu wurden gleichzeitig 3.700 erwachsene Teilnehmer des TNS-Online-Panels<sup>5)</sup> eingeladen, eine inhaltlich identische Befragung am Bildschirm auszufüllen. Hieraus liegen zusätzlich 1.049 Antworten vor (im Folgenden: „online“-Befragung). Ziel dieser Ergänzungsbefragung war zu prüfen, ob eine derartige internetbasierte Panel-Befragung hinreichend ähnliche Ergebnisse erbringt, so dass künftig bei vergleichbaren Bewertungsprojekten auf dieses wesentlich kostengünstigere Verfahren gesetzt werden könnte.

### 3. DATENANALYSE UND ERGEBNISSE

#### 3.1. Aktuelle Erholungsleistung: Walderholung im Status Quo

##### 3.1.1. Waldbesucheranteil

Zunächst stand in Frage, welcher Anteil der Bevölkerung in Deutschland überhaupt Waldbesuche unternimmt. Aus der CAPI-Befragung geht hervor, dass gut drei Viertel (76,4%)<sup>6)</sup> aller Befragten während der letzten zwölf Monate in ihrer Freizeit mindestens einmal

einen Wald in der Nähe ihres jeweiligen Wohnortes besucht haben (einschließlich Wochenendausflügen, ohne Urlaubsreisen), die übrigen seltener oder gar nicht. Zählt man letztere mit, so überwiegen in der Bevölkerung die sporadischen Waldbesucher: Etwa die Hälfte aller Befragten besuchen den Wald dreimal pro Jahr oder seltener (48,9%); etwa zwei Drittel besuchen ihn seltener als zwölfmal pro Jahr (67,3%). Rechnerisch ergibt sich eine mittlere Besuchsfrequenz aller Befragten von 27,5 Waldbesuchen pro Jahr, bei einer Standardabweichung von  $sd=58,9$ .<sup>7)</sup> Die Häufigkeitsverteilung der Besuche geht aus *Abbildung 2* hervor.

Im Vergleich zur CAPI-Umfrage zeigen die online gewonnenen Ergebnisse zwar ein ähnliches Muster, im Detail aber auch deutliche Unterschiede. Insbesondere ist der Anteil der Waldbesucher hier um zehn Prozentpunkte höher (86,4%)<sup>8)</sup>, und der rechnerische Durchschnitt der Besuchshäufigkeit beträgt 37,4 Besuche pro Jahr ( $sd=70,4$ ). Nach dem KOMOLGOROV-SMIRNOV-Test unterscheiden sich die beiden Häufigkeitsverteilungen hochsignifikant – zumindest dann, wenn man sämtliche Befragten vergleicht ( $\alpha > 0,1\%$ ). Werden lediglich die Waldbesucher verglichen, so sind die Unterschiede geringer, und das Testergebnis ist nahe der Signifikanzgrenze ( $\alpha = 9,1\%$ ).

<sup>5)</sup> Das TNS-Online-Panel besteht aus etwa 80.000 zufällig offline rekrutierten Personen, die ihre Internet-Kontaktdaten den am Panel beteiligten Meinungsforschungsunternehmen zur Verfügung gestellt und sich bereit erklärt haben, sporadische Einladungen zu (unentgeltlichen) Befragungen zu akzeptieren.

<sup>6)</sup> Die 5%-Konfidenzgrenzen betragen 73,6% und 78,9% (Berechnung nach HALD, 1952).

<sup>7)</sup> Für diese Berechnung wurde jeder Besuchshäufigkeitsklasse die jeweilige Klassenmitte unterstellt, der obersten Besuchshäufigkeitsklasse „über 365 Besuche/Jahr“ eine Besuchshäufigkeit von 400 Besuchen/Jahr.

<sup>8)</sup> 5%-Konfidenzgrenzen: 84,4% und 88,3% (Berechnung wie oben).

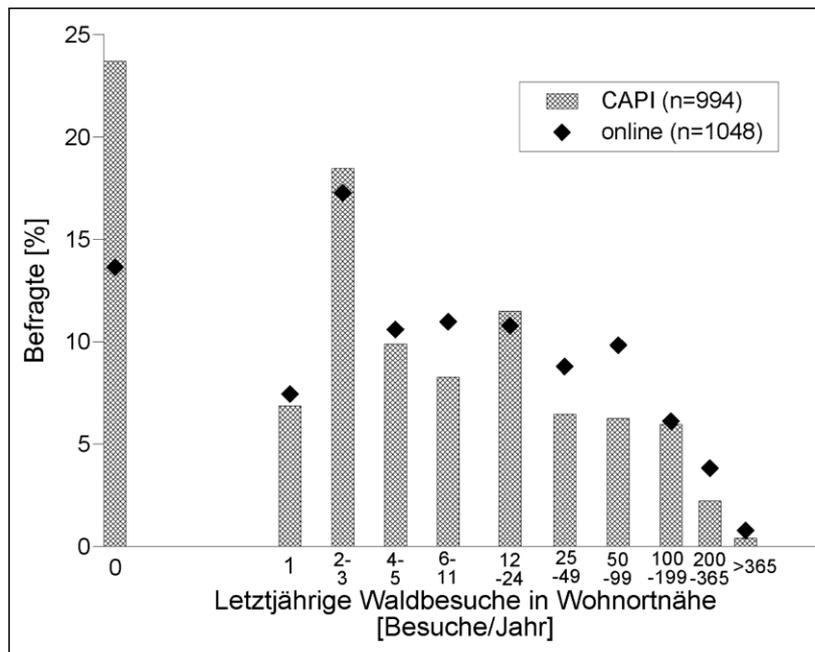


Abb. 2

Häufigkeitsverteilung der jährlichen Waldbesuche.

Frequency distribution of annual forest visits.

### 3.1.2. Durchschnittliche Zahlungsbereitschaften

Ihre Zahlungsbereitschaft für Waldbesuche im Status Quo haben in der CAPI-Umfrage insgesamt 596 Interviewte angegeben; bezogen auf die Nettostichprobe beträgt die fragespezifische Antwortverweigerung hier also 41,0%. Wie zuvor erweist es sich auch bei den Zahlungsbereitschaften als zweckmäßig, zwischen sämtlichen Befragten einerseits und Waldbesuchern andererseits zu unterscheiden. Im Durchschnitt sämtlicher Befragten ergibt sich eine Zahlungsbereitschaft von 26,94 € pro Person und Jahr; diejenige der Waldbesucher allein beträgt 32,06 €/p/a. *Tabelle 1* zeigt dieses Ergebnis sowie weitere Maße der Verteilung der Zahlungsbereitschaften.

Erneut sind die Antworten aus der online-Umfrage ähnlich, aber nicht gleich. Insbesondere fällt auf, dass die fragespezifische Verweigerung hier mit 13,4% deutlich geringer ist als zuvor. Die Zahlungsbereitschaften selbst unterscheiden sich dagegen weniger. Der entsprechende Mittelwert für alle Befragten beträgt 30,32 €/p/a, derjenige der Waldbesucher allein 32,38 €/p/a. Ähnlich wie zuvor unterscheiden sich die Ergebnisse von CAPI- und online-Umfrage statistisch nur dann, wenn auch die Nicht-Besucher eingeschlossen

werden; betrachtet man allein die Zahlungsbereitschaft der Waldbesucher, so gibt es keine signifikanten Unterschiede zwischen den beiden Befragungen (nach dem MWU-Test betragen die Irrtumswahrscheinlichkeiten für alle Befragten  $\alpha=0,02\%$ , für die Waldbesucher  $\alpha=85,9\%$ ).

### 3.1.3. Hochrechnung auf die Wälder der Bundesrepublik Deutschland

Mit Hilfe der allgemeinen Bevölkerungsstatistik kann die Zahlungsbereitschaft nun auf die Bundesrepublik hochgerechnet werden. Nach den aktuellsten verfügbaren Daten betrug Ende 2009 der Bevölkerungsstand Deutschlands (ab 14 Jahren) 71,5 Mio. Personen (STBA, 2011:42). Nach der CAPI-Umfrage sind davon knapp 55 Millionen Waldbesucher; für sie ergibt sich eine aggregierte Zahlungsbereitschaft i.H.v. knapp 1,8 Mrd. €/Jahr. Berücksichtigt man zusätzlich die Zahlungsbereitschaft der Nicht-Besucher, so beträgt die Zahlungsbereitschaft der Gesamtbevölkerung über 14 Jahre für Waldbesuche 1,9 Mrd. €/Jahr. Auf Basis der online-Umfrage ergeben sich, ebenfalls auf die Bevölkerung ab 14 Jahren bezogen, jeweils geringfügig höhere Schätzungen (s. *Tabelle 2*).

Tab. 1  
Zahlungsbereitschaften für Walderholung im Status Quo.  
WTP for forest recreation in the status quo.

Datensatz	gültige Antworten [n]	Mittelwert [€/p/a]	sd [€/p/a]	Median [€/p/a]	Min [€/p/a]	Max [€/p/a]	Anzahl Null-ZB** [n]
CAPI	596	26,94	59,2	10	0	1000	167
davon: nur Waldbesucher*	453	32,06	63,2	20	0	1000	78
online	908	30,32	59,6	15	0	1000	92
davon: nur Waldbesucher*	784	32,38	62,5	15	0	1000	57
*mind. 1 Waldbesuch in der Freizeit in den letzten 12 Monaten, einschl. Wochenendausflüge, ohne Urlaubsreisen							
** nur echte Null-Zahlungsbereitschaft (ohne Protestvoten)							

Tab. 2  
Hochrechnung der mittleren Zahlungsbereitschaften auf die Gesamtbevölkerung.  
Projection of the total population's aggregated WTP.

	[n]	ZB (nach CAPI)		ZB (online)	
		Mittel [€/p/a]	aggregiert [T€/a]	Mittel [€/p/a]	aggregiert [T€/a]
Bevölkerungsstand 31.12.2009 (ab 14 Jahre)	71.559.078	26,94	1.928.016	30,32	2.169.313
davon Waldbesucher: 76,36% (nach CAPI)	54.642.512	32,06	1.752.058	-	-
davon Waldbesucher: 86,35% (online)	61.791.264	-	-	32,38	2.000.888

Tab. 3

**Zahlungsbereitschaften für Walderholung im Status Quo nach Bundesländern  
(gepoolt aus CAPI- und online-Umfrage).**

**WTP for forest recreation in the status quo by federal states (CAPI and online data pooled).**

Land	alle Befragten			davon: nur Waldbesucher		
	ZB	sd	n	ZB	sd	n
Schleswig-Holstein	30,86	52,6	50	36,72	56,3	36
Hamburg	24,83	30,1	47	32,73	33,8	30
Niedersachsen	26,69	49,8	153	32,16	55,8	115
Bremen	11,45	14,6	11	12,63	17,2	8
NRW	28,60	64,5	340	32,42	69,5	285
Hessen	28,47	51,6	116	30,85	54,5	101
Rheinland-Pfalz	42,40	73,7	78	46,15	77,7	67
Baden-Württemberg	33,81	60,8	182	37,29	65,2	152
Bayern	33,39	84,4	194	36,22	89,1	172
Saarland	18,96	26,7	28	22,08	27,7	24
Berlin	20,47	40,7	64	19,93	20,8	46
Brandenburg	19,22	26,5	51	21,20	27,5	45
Mecklenburg-Vp.	26,03	53,0	29	30,88	51,0	17
Sachsen	29,73	54,3	71	26,34	32,0	64
Sachsen-Anhalt	25,35	26,9	48	26,60	27,9	42
Thüringen	25,35	35,6	42	22,27	38,7	33
<b>Bundesrepublik</b>	<b>28,98</b>	<b>59,4</b>	<b>1504</b>	<b>32,27</b>	<b>62,8</b>	<b>1237</b>

### 3.1.4. Regionale Unterschiede der Zahlungsbereitschaft

Da in der Befragung auch der Herkunftsort der Interviewpartner erhoben wurde, lassen sich einige Aussagen zur regionalen Verteilung der Zahlungsbereitschaft treffen. Für die in *Tabelle 3* gezeigte Auswertung nach Bundesländern wurden die Ergebnisse beider Umfragen gepoolt, da die Stichprobenumfänge in den einzelnen Ländern teilweise sehr gering sind; die Tabelle sollte entsprechend behutsam interpretiert werden.

Tendenziell zeichnet sich ab, dass überdurchschnittliche Zahlungsbereitschaften in den meisten westlichen Flächenländern auftreten; Rheinland-Pfalz nimmt hierbei sowohl bei den Waldbesuchern als auch bei allen Befragten zusammen die Spitzenposition ein. Vergleichsweise geringere Zahlungsbereitschaften haben die Befragten in den meisten östlichen Bundesländern sowie in den Stadtstaaten Berlin und dem sehr gering bewaldeten Bremen, welches das Schlusslicht bildet. Ob solche augenscheinlichen regionalen Mittelwertdifferenzen aber auf eine unterschiedliche Wertschätzung von Wäldern in den jeweiligen Ländern auch unter ceteris-paribus-Bedingungen zurückzuführen sind, oder ob sie schlicht an regional unterschiedlicher Besucherzusammensetzung liegen, lässt sich durch einen einfachen Mittelwertvergleich nicht beurteilen. Einen besseren Aufschluss ergibt die multiple Regression der Zahlungsbereitschaften.

### 3.1.5. Regressionsanalyse der Zahlungsbereitschaft im Status Quo

Diese Regressionsanalyse basierte auf dem Grundmodell  $ZB = \beta_0 + \beta_n X_n + \varepsilon$ , bei dem  $X$  ein Vektor von

Erklärungsvariablen ist;  $\beta$  sind die zu schätzenden Koeffizienten, und  $\varepsilon$  die verbleibenden Residuen. Alle Schätzungen erfolgten über die Methode der kleinsten Quadrate (OLS). In die Ausgangsschätzungen gingen folgende kardinal skalierte Erklärungsvariablen ein: die Anzahl der letztjährigen Waldbesuche in Wohnortnähe; die Entfernung zum zuletzt besuchten Wald; das individuelle monatliche Äquivalenzeinkommen<sup>9)</sup> des jeweiligen Befragten; sein bzw. ihr Alter. Zusätzliche Charakteristika wurden als Dummy-Variablen aufgenommen, nämlich der Befragungsmodus (2 Kategorien: CAPI oder online), das Geschlecht des Befragten (2 Kategorien), die Einwohnerzahl seines Herkunftsortes (7 BIK-Ortsgrößenklassen), das entsprechende Bundesland (16 Länder), seine Berufsgruppenzugehörigkeit (zusammengefasst zu 7 Gruppen: in Ausbildung, Arbeiter, Angestellte, Beamte, Selbständige, Rentner, ohne Erwerbseinkommen) sowie die beim letzten Waldbesuch hauptsächlich ausgeübte Aktivität (zusammengefasst zu 7 Kategorien: Spazierengehen/Wandern, Radfahren, sonstiger Sport

<sup>9)</sup> Das Äquivalenzeinkommen ist ein Rechenmaß, mit dem das individuell verfügbare Einkommen von Personen aus unterschiedlich großen Haushalten miteinander vergleichbar gemacht wird. Skaleneffekte (durch Zunahme des individuell verfügbaren Einkommens bei zunehmender Haushaltsgröße) werden dabei über eine entsprechende Gewichtung eliminiert. Die Berechnung erfolgte nach der „modifizierten OECD-Skala“ (HAGENAARS *et al.*, 1994). Da das Haushaltseinkommen aus der Befragung nur klassiert vorlag, wurde es zuvor nach dem bei HSIAO 1983 beschriebenen Verfahren kardinalisiert (Anpassung einer Log-Normalverteilung an die empirische Einkommensverteilung; Bestimmung der jeweiligen Klassenmitten anschließend aus dieser Verteilung).

einschließlich Reiten, Hundeausführen, Naturbeobachtung und Fotografieren, Beeren und Pilzesammeln, Sonstiges einschließlich Jagen und Fischen). Das Grundmodell wurde in linearer, semilogarithmischer und doppelt-logarithmischer Formulierung getestet. Ausgehend vom vollständigen Satz der Erklärungsvariablen wurden diese einzeln auf Signifikanz geprüft und bei einem  $\alpha$ -Fehler über 10% schrittweise rückwärts eliminiert; lediglich die Beibehaltung der Variablen „Befragungsmodus“ wurde erzwungen, um die Unterschiede zwischen CAPI- und Online-Befragung explizit beziffern zu können. In Ergänzung zum beschriebenen Grundmodell wurde auch ein weiteres Modell mit demselben Verfahren überprüft, in das als abhängige Variable anstelle der Zahlungsbereitschaft pro Jahr die Zahlungsbereitschaft pro Besuch einging.

Da die Zahlungsbereitschaften angenähert log-normalverteilt sind, ergaben sich befriedigende Residuenverteilungen in allen geprüften Fällen nur dann, wenn die jeweilige abhängige Variable logarithmiert worden war. *Tabelle 4* zeigt die Ergebnisse der Doppellogmodelle nach Eliminierung der insignifikanten Erklärungsvariablen.<sup>10)</sup>

<sup>10)</sup> Die Semilogmodelle zeigen strukturell sehr ähnliche Ergebnisse, jeweils aber bei niedrigerem  $r^2$ .

Die Ergebnisse sind durchgehend mit den ökonomischen Erwartungen konsistent: Mit steigendem Einkommen steigt stets auch die Zahlungsbereitschaft für Waldbesuche (aufgrund abnehmender Wirksamkeit der Budgetrestriktionen), und eine höhere Besuchshäufigkeit steigert die Zahlungsbereitschaft pro Jahr (aufgrund der höheren Konsummenge), senkt aber die Zahlungsbereitschaft pro Besuch (aufgrund sinkender Grenznutzen zusätzlicher Besuche). Die zugehörigen Koeffizienten sind sämtlich statistisch hochsignifikant von Null verschieden. Die übrigen Variablen sind deskriptiv zu interpretieren in dem Sinne, dass sich zu ihnen aus ökonomischer Sicht keine eindeutigen Vorhersagen treffen lassen. Die Konstante ist in keinem der beiden dargestellten Modelle signifikant. Zunehmende Entfernung zum besuchten Wald steigert in beiden Fällen die Zahlungsbereitschaft. Hier schlägt sich vermutlich nieder, dass die Inkaufnahme einer weiteren Anreise tendenziell eine höhere Wertschätzung reflektiert. Bei den weiteren Variablen lässt sich kein durchgängiges Muster erkennen. Männer haben zwar eine geringere jährliche Zahlungsbereitschaft als Frauen – bei der Zahlungsbereitschaft pro Besuch zeigt sich dieser Unterschied dagegen nicht. Nur wenige Bundesländer unterscheiden sich, je nach Modell, zumindest schwach signifikant von den übrigen; hochsignifikant ist dies

Tab. 4

**Regressionsanalysen der Zahlungsbereitschaft: Ergebnisse des doppelt logarithmischen Schätzansatzes nach Elimination insignifikanter Erklärungsvariablen.**

**Regression analyses of WTP: Results of the double log approach after stepwise backwards elimination of insignificant explanatory variables.**

Variable	abhängig: ln (ZB)		abhängig: ln (ZB/Besuch)		
	$\beta$	t	$\beta$	t	
Konstante	-0,58625	0,926	0,13306	0,237	
Besuchshäufigkeit	ln [Bes./Jahr]	0,26406	9,557***	-0,81192	33,06***
Entfernung	ln [km]	0,11860	2,921***	0,07561	2,108**
Äquivalenzeinkommen	ln [€/m/p]	0,34385	3,959***	0,32804	4,268***
Befragungsmodus	online	0,07595	0,839	-0,19204	2,380**
Geschlecht	männlich	-0,19469	2,270**	-	-
Berufsgruppe	selbständig	-	-	-0,24456	1,764*
Besuchsaktivität	Sonstiges	0,81764	2,542**	-	-
Land	Brandenburg	-	-	-0,36847	1,852*
Land	Bremen	-1,78227	2,913***	-	-
Land	Thüringen	-0,44499	1,801*	-	-
BIK-Ortsgrößenklasse	50-100 Tsd.EW	-	-	0,22957	1,828*
$r^2$	[%]	12,332		59,392	
$r^2$ (korrigiert)	[%]	11,626		59,072	
F-Test		17,46***		185,54***	
Freiheitsgrade		8+993		7+888	
Durbin-Watson-Test		1,98157		2,04281	

(kursiv: |t|-Werte; \*: signifikant mit  $\alpha < 10\%$ ; \*\*:  $\alpha < 5\%$ ; \*\*\*:  $\alpha < 1\%$ ).

lediglich in Bremen, und auch nur in einem der beiden Modelle. Auch die jeweiligen Waldbesuchsaktivitäten haben keinen durchgängigen Einfluss auf die Zahlungsbereitschaft: Mit einer Ausnahme wurden sämtliche der entsprechenden Dummies als insignifikant eliminiert. Entsprechendes trifft auf Ortsgrößenklasse und Berufsgruppe der Befragten zu.

Zusammenfassend ergibt sich also: Die oben gezeigten regionalen Mittelwertunterschiede gehen nach der Regressionsanalyse offensichtlich nicht darauf zurück, dass regionale Eigenschaften des Waldes unterschiedliche Wertschätzungen bewirken würden. Vielmehr schlägt sich hier die regional unterschiedliche Besucherzusammensetzung nieder (namentlich hinsichtlich Einkommen und Besuchshäufigkeit); darüberhinausgehende regionale und soziodemographische Einflüsse sowie auch der Einfluss der jeweiligen Waldbesuchsaktivitäten sind aber nicht eindeutig nachweisbar. Schließlich zeigt sich hier erneut, dass der Befragungsmodus einen gewissen Einfluss auf die ermittelte Zahlungsbereitschaft hat, aber keinen sehr dominanten: Bei der Zahlungsbereitschaft pro Jahr sind die Unterschiede sehr klein und nicht signifikant von Null verschieden; die pro Besuch umgerechnete Zahlungsbereitschaft ist in der online-Befragung mit  $\alpha < 5\%$  niedriger als in der CAPI-Befragung.

### 3.2. Potentielle Erholungsleistung

#### 3.2.1. Verbesserungswünsche der Befragten

An das Leistungsspektrum der Wälder werden sowohl seitens der allgemeinen Bevölkerung als auch seitens der Forstbetriebe unterschiedliche und teilweise konfligierende Ansprüche gestellt; auch das weitverbreitete Leitbild einer multifunktionalen Bewirtschaftung (MARWEDE *et al.*, 2007) erfordert grundsätzlich Kompromisse. Entsprechend ist zu erwarten, dass aus individueller Sicht Verbesserungspotentiale bestehen (welche nicht unbedingt wechselseitig kompatibel sein müssen). Ein

Ziel der hier dargestellten Befragung war, solche Verbesserungswünsche sowie ihren potentiellen Einfluss auf den Erholungsnutzen in Erfahrung zu bringen.

Beim Blick auf die von den Befragten genannten Verbesserungswünsche fällt zunächst ins Auge, dass sich die Antworten relativ gleichmäßig über die im Fragebogen angebotenen Items verteilen; lediglich der Wunsch nach weniger Müll bzw. aufgeräumteren Wäldern sticht hervor, den sowohl in der CAPI- als auch in der online-Umfrage jeweils über ein Viertel der Befragten als wichtigstes Anliegen nannten. *Abbildung 3* zeigt die Häufigkeitsverteilung der Antworten im Einzelnen.

Sie lassen sich zwanglos zu drei Gruppen zusammenfassen: „Naturschutzorientierte“ Anliegen umfassen den generellen Wunsch nach ‚größerer Naturnähe‘ der Wälder, nach ‚mehr Laub- und Mischwaldanteilen‘, nach ‚höherer Strukturvielfalt‘ und nach einem Verzicht auf Holzfällung bzw. Maschineneinsatz (letzteres entstammt den offen abgefragten ‚sonstigen‘ Anliegen). Eine zweite Gruppe kann mit ‚infrastrukturorientierten‘ Anliegen bezeichnet werden (bestehend aus dem bereits oben erwähnten Wunsch nach ‚weniger Müll bzw. aufgeräumteren Wäldern‘, nach ‚mehr oder besseren Wegen‘, nach ‚mehr oder besseren Erholungsangeboten‘, nach ‚verbessertes Sicherheit‘ sowie auch nach ‚weniger Störungen durch andere Waldbesucher‘). Eine dritte Gruppe bilden schließlich die Antworten derjenigen Befragten, die entweder explizit keine Verbesserungswünsche hatten oder hierzu keine Angaben machten (ein geringer Rest der Antworten entfällt auf ‚Sonstiges‘). *Tabelle 5* zeigt die prozentualen Verteilungen dieser gruppierten Anliegen.

Etwa die Hälfte der Wünsche bezieht sich auf infrastrukturelle Verbesserungen – sowohl bei der Frage nach dem wichtigsten als auch nach den drei wichtigsten Anliegen der Befragten. CAPI- und online-Befragung unterscheiden sich hier nicht wesentlich. Anders ist dies bei den ‚naturschutzorientierten‘ Anliegen, die jeweils knapp ein Viertel der persönlich Befragten, aber

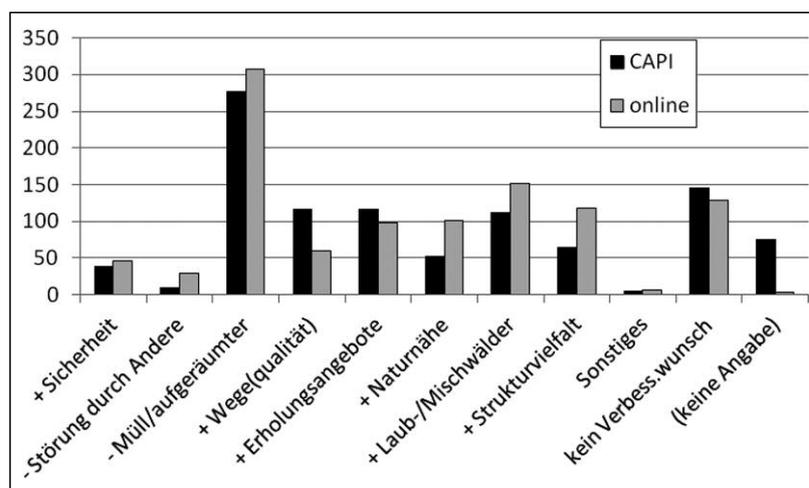


Abb. 3

Wichtigster Verbesserungswunsch der Befragten für den Wald  
(Häufigkeitsverteilung über alle angebotenen Items).

Most important improvement request  
(frequency distribution of all offered items).

Tab. 5

**Verbesserungswünsche der Befragten, gruppiert nach Art der Anliegen (in % der Antworten).**  
**Improvement requests of the interview partners, grouped by type of request (in % of answers).**

	„naturschutzorientierte“ Anliegen	„infrastrukturorientierte“ Anliegen	keine Anliegen	(Sonstiges)
wichtigstes Anliegen				
CAPI (n=1.011)	22,8	55,1	21,8	0,3
online (n=1.049)	35,5	51,6	12,5	0,4
3 wichtigste Anliegen				
CAPI (n=3.033)	23,3	48,2	28,4	0,1
online (n=3.147)	36,8	46,6	16,3	0,3

über ein Drittel der online Befragten äußerten. Bemerkenswerterweise nannten zudem viele Interviewpartner überhaupt keine Verbesserungswünsche für die Waldbewirtschaftung. In der persönlichen Befragung waren dies mehr als ein Fünftel; online waren es weniger, was überwiegend auf den hier sehr geringen Anteil fehlender Angaben zurückzuführen ist (dies ging bereits aus *Abbildung 3* hervor).

Aus ökonomischer Sicht ist zu fragen, inwieweit die von den Interviewpartnern geäußerten Präferenzen für Verbesserungen der Waldwirtschaft deutlich genug ausgeprägt sind, um gegebenenfalls auch Änderungen des beobachtbaren Verhaltens zu bewirken, und inwieweit dies gegen möglichen „cheap talk“ (FARRELL und RABIN, 1996) abgegrenzt werden kann. Einen ersten Aufschluss hierzu geben die Auskünfte zu der Frage, ob die Interviewpartner nach einer Umsetzung ihrer jeweiligen Verbesserungswünsche den Wald häufiger besuchen würden oder nicht. Tatsächlich verneinte dies die überwiegende Mehrheit selbst derjenigen Befragten, welche zuvor Verbesserungswünsche geäußert hatten (79,8% in der CAPI-Umfrage; in der online-Umfrage war der Anteil der Nein-Antworten geringer, betrug mit 56,3% aber immer noch mehr als die Hälfte). Bereits dieser Befund lässt vermuten, dass der ökonomische Nutzen einer Umsetzung einiger der Verbesserungsanliegen insgesamt gering sein könnte. Bei Betrachtung der entsprechenden zusätzlichen Zahlungsbereitschaft wird dies noch deutlicher.

### 3.2.2. Zusätzliche Zahlungsbereitschaft bei veränderter Waldbehandlung

Für die von ihnen gewünschte wichtigste Verbesserung äußerten lediglich 7,9% aller Befragten der CAPI-Umfrage eine gültige zusätzliche Zahlungsbereitschaft größer Null (online waren es immerhin 20 %).<sup>11)</sup> Die gültigen Antworten betragen also zu 92% (bzw. 80%) Null. Rechnet man die Null-Zahlungsbereitschaften mit ein, so beträgt der Mittelwert für die jeweils individuell wichtigste Verbesserung über alle Befragten in der CAPI-Umfrage mit 96 Cent weniger als 1 € pro Person und Jahr (online waren es 3,92 €/p/a).<sup>12)</sup>

Selbst wenn man ausschließlich die positiven Zahlungsbereitschaften – ohne Null – betrachtet, beträgt der Mittelwert lediglich 12,14 (online: 19,39) €/p/a. Diese Mittelwerte beruhen auf den Angaben von nur 35 (bzw. 164) Befragten mit positiver zusätzlicher Zahlungsbereitschaft. Sie sind zudem nahezu gleich über die beiden oben abgegrenzten Gruppen von Anliegen verteilt: In der persönlichen Befragung nannten 17 Befragte durchschnittlich zusätzliche 13,12 €/p/a für einen naturschutzbezogenen Verbesserungswunsch; für die übrigen 18 war eine infrastrukturelle Verbesserung das wichtigste Anliegen, für das sie zusätzlich 11,22 €/p/a zu zahlen bereit waren (in der online-Befragung betragen die entsprechenden Gruppengrößen 60 zu 102,<sup>13)</sup> und die zugehörigen Zahlungsbereitschaften 17,02 zu 21,02 €/p/a; hier überwogen also zahlenmäßig die Befürworter infrastruktureller Anliegen, und auch ihre Zahlungsbereitschaft war geringfügig höher).

Das Bild verändert sich kaum, wenn man die zusätzliche Zahlungsbereitschaft für alle drei Verbesserungen betrachtet, welche den Befragten am Herzen lagen. Bei Einschluss der Null-Werte beträgt sie in der CAPI-Umfrage im Mittel 1,19 €/p/a (online: 5,70 €/p/a) und

<sup>11)</sup> Aus befragungstechnischen Gründen und zur Vermeidung von Missverständnissen wurden die Interviewpartner explizit nicht nach ihrer zusätzlichen Zahlungsbereitschaft nach Umsetzung ihrer individuellen Verbesserungswünsche gefragt, sondern nach dem entsprechenden jährlichen Gesamtbetrag. Dieser setzt sich aus der Zahlungsbereitschaft für den Status Quo und der zusätzlichen Zahlungsbereitschaft zusammen.

<sup>12)</sup> Im Folgenden sind die Unterschiede zwischen CAPI- und online-Befragung durchgehend im statistischen Sinn signifikant. Von den jeweiligen Effektgrößen her (insbesondere den Unterschieden zwischen den jeweiligen Zahlungsbereitschaften) sind die absoluten Differenzen aber relativ gering; zudem haben viele Interviewpartner die Angabe ihrer zusätzlichen Zahlungsbereitschaft verweigert (dies bewirkt reduzierte Fallzahlen sowie einen zusätzlichen Selbstselektionseffekt). Auf Darstellung der Ergebnisse der formalen Signifikanztests wird daher verzichtet.

<sup>13)</sup> Bei den zwei zu 164 fehlenden Beobachtungen handelt es sich um „Sonstige“ Nennungen, die weder naturschutz- noch infrastrukturellen Anliegen zugeordnet werden konnten.

Tab. 6

**Zusätzliche Zahlungsbereitschaften für Walderholung  
bei Einführung gewünschter Verbesserungen.**

**Marginal WTP for forest recreation if required improvements were implemented.**

<i>Datensatz</i>	<i>gültige Antworten [n]</i>	<i>Mittelwert [€/p/a]</i>	<i>sd [€/p/a]</i>	<i>Median [€/p/a]</i>	<i>Min [€/p/a]</i>	<i>Max [€/p/a]</i>	<i>Anzahl 0-ZB [n]</i>
<b>ZB für wichtigstes Anliegen</b>							
CAPI	444	0,96	4,2	0	0	40	409
online	811	3,92	16,0	0	0	300	647
<b>ZB für 3 wichtigste Anliegen</b>							
CAPI	443	1,19	5,1	0	0	50	409
online	802	5,70	22,3	0	0	350	647

übertrifft damit die vorgenannten Werte für die wichtigste Verbesserung jeweils nur leicht.<sup>14)</sup>

Addiert man die angegebenen zusätzlichen Zahlungsbereitschaften zu der jeweiligen Basis-Zahlungsbereitschaft im Status Quo, so ergeben sich nur recht geringfügige Änderungen: Selbst eine Einführung sämtlicher gewünschter Verbesserungen – wenn diese denn überhaupt möglich wäre – erhöhte die durchschnittliche Zahlungsbereitschaft für Waldbesuche (vgl. *Tabelle 1*) lediglich auf 28,13 €/p/a nach den Ergebnissen der CAPI-Umfrage (bzw. auf 36,01 €/p/a nach den online gewonnenen Ergebnissen). Die auf die Gesamtbevölkerung hochgerechneten Werte (vgl. *Tabelle 2*) stiegen um 4,4% auf 2,0 Mrd. €/a nach der CAPI-Befragung (bzw. um 18,8% auf 2,6 Mrd. €/a online).

Bei der Interpretation dieser Zahlen muss berücksichtigt werden, dass die von den Befragten gewünschten Verbesserungen sich zumindest teilweise gegenseitig ausschließen und daher nicht durchgehend auf denselben Flächen realisierbar sind. So ist z.B. ein Ausbau weiterer Erholungsangebote tendenziell nicht mit einer Vergrößerung der Naturnähe vereinbar. Daher stellen die hier ermittelten zusätzlichen Zahlungsbereitschaften Obergrenzen für das Potential dar, den Erholungsnutzen der Wälder durch veränderte Bewirtschaftung zu steigern, welche durch konkrete Veränderungen der Waldbewirtschaftung nicht vollständig ausschöpfbar sind.

## 4. DISKUSSION

### 4.1. Zur Bewertungsmethode

Mit dieser Studie werden aktuelle Schätzungen des ökonomischen Wertes der Erholung im Wald vorgelegt, welche hinsichtlich des verwendeten Bewertungsinstruments (Bewertungsmethode und ihre Umsetzung in konkrete Fragebogenformulierungen) auf entsprechende Studien aus den 1990er Jahren aufbauen. Zur grund-

sätzlichen Methodendiskussion sei zunächst auf die damalige Literatur verwiesen (insbesondere die umfangreichen Studien von LÖWENSTEIN, 1994 und ELSASSER, 1996).

Für den Rückgriff auf dieses bereits entwickelte Bewertungsinstrument in der aktuellen Studie sprach – neben den dadurch eröffneten Vergleichsmöglichkeiten – in erster Linie, dass seine Anfälligkeit gegenüber diversen möglichen Fehlerquellen damals ausgiebig empirisch getestet und weitgehend eingegrenzt worden ist (konkret u.a. Tests auf strategisches Verhalten, Verankerungsanreize, Zuordnungsfehler, den Einfluss des Bezahlungskartenformats und der Revisionsmöglichkeit für die Zahlungsbereitschaft). Dadurch lässt sich auch recht gut einschätzen, wie robust mit diesem Konzept gewonnene Ergebnisse sind. Nach den damaligen Experimenten verzerrten die einzelnen geprüften Fehlerquellen die ermittelten Zahlungsbereitschaften unter Testbedingungen um bis zu etwa 60% oder weniger. Bei summarischer Betrachtung aller geprüften Fehlerquellen zeigte sich, dass das Instrument insgesamt relativ „vorsichtige“ Bewertungen der Erholungsleistung liefert (ELSASSER, 1999:180) und auch externen Validitätstests (durch entsprechende Reisekostenanalysen) standhält.

Diese Ergebnisse sind nicht unmittelbar auf die Ermittlung der zusätzlichen Zahlungsbereitschaft bei Umsetzung der Verbesserungswünsche der Befragten in der aktuellen Studie übertragbar, da die damaligen Untersuchungen ausschließlich auf die Zahlungsbereitschaft im Status Quo abzielten. Allerdings wurde die Ermittlung der zusätzlichen Zahlungsbereitschaft eng an das bereits bestehende Bewertungskonzept angelehnt. Soweit die erwähnten Validitätsprüfungen übertragbar sind, ist daher auch von einer eher „vorsichtigen“ Bewertung der zusätzlichen Zahlungsbereitschaft auszugehen.

### 4.2. Vergleich zwischen CAPI- und Online-Befragung

Die CAPI-Befragung beruhte auf einer separat gezogenen Stichprobe der Gesamtbevölkerung; die Befragten

<sup>14)</sup> Eine eindeutige Zuordnung dieser Zahlungsbereitschaften zu naturschutz- bzw. infrastrukturorientierten Anliegen ist hier nicht mehr möglich, da die „drei wichtigsten“ Anliegen oft aus beiden Gruppen stammen.

wurden persönlich in ihren Haushalten interviewt. Demgegenüber wurde für die Online-Befragung auf ein Panel zurückgegriffen, welches zwar ebenfalls offline rekrutiert worden ist, aber nur Internetnutzer enthält; diese beantworteten die Fragen an ihrem Bildschirm.

Im Vergleich der beiden Befragungsmodi gilt die persönliche Befragung einer repräsentativen Bevölkerungstichprobe zwar grundsätzlich als das überlegene Verfahren. Aber auch sie bietet keine unstrittige Referenz, weil die Stichprobenziehung hier ebenfalls durch Selbstselektion der Befragten mitbeeinflusst wird, welche sich in beschränkter Stichprobenausschöpfung niederschlägt (hier: 40%). In diesem Zusammenhang sind darüber hinaus die Ausfallquoten bei den Einzelfragen von Belang. In der vorliegenden online-Befragung waren diese teilweise deutlich geringer als in der persönlichen Befragung.

Bei den jeweiligen empirischen Ergebnissen zeigten sich keine fundamentalen Unterschiede: CAPI- und online-Befragung erbrachten durchgehend ähnliche, wenn auch nicht gleiche Resultate. Bei der Untersuchung der Erholungsleistung im Status Quo divergieren die Ergebnisse der beiden Befragungsmodi vor allem hinsichtlich der jeweils geschätzten Waldbesucheranteile und der entsprechenden Besuchshäufigkeiten. Dagegen unterscheiden sich die ermittelten Zahlungsbereitschaften der Gesamtbevölkerung weniger, und die der Waldbesucher gar nicht (d.h. die Unterschiede sind nicht signifikant von Null verschieden). Auch die Untersuchung der potentiellen Erholungsleistung erbrachte strukturell sehr ähnliche Ergebnisse über die Verbesserungswünsche der Befragten; die zusätzlichen Zahlungsbereitschaften unterscheiden sich wieder etwas deutlicher, sind aber im Vergleich zum Status Quo ohnehin gering und fußen hier bei beiden Befragungsmodi auch auf geringeren Fallzahlen. Die jeweiligen Größenordnungen der Zahlungsbereitschaft stimmen bei allen hier untersuchten Fragen überein.

In der Summe erscheinen die online gewonnenen Ergebnisse daher nicht als grundsätzlich weniger vertrauenswürdig im Vergleich zu den CAPI-Daten. Ähnliche Einschätzungen überwiegen auch in der Literatur über diesbezügliche Vergleichsexperimente (vgl. LINDHJEM und NAVRUD, 2011 für eine aktuelle Übersicht).

### 4.3. Vergleich mit früheren Schätzungen des Erholungswertes für Alltagsbesucher

Im Vergleich mit den früheren Schätzungen ist die hier ermittelte Zahlungsbereitschaft nach beiden Befragungsmodi niedriger – selbst wenn man nur nominale Werte vergleicht, d.h. die zwischenzeitliche Geldentwertung außer Betracht lässt. Für Mitte der 1990er Jahre kam eine Metaanalyse der bis dato vorliegenden Studien für die Bundesrepublik je nach Rechenvariante auf aggregierte Werte der Erholungsleistung für Alltagsbesucher zwischen 2,5 und 2,9 Mrd. €/a, bezogen auf die damalige Bevölkerung über 14 Jahren mit mindestens einem jährlichen Waldbesuch (ELSASSER, 2001). Umgerechnet auf den aktuellen Geldwert wären dies zwischen 3,5 und 4,1 Mrd. €/a.<sup>15)</sup> Die heutigen Schätzungen für eine gleich abgegrenzte Grundgesamtheit liegen laut

Tabelle 2 zwischen 1,8 und 2,2 Mrd. €/a, sind also deutlich geringer.

Kann daraus geschlossen werden, dass sich die Präferenzen in der Gesellschaft tatsächlich verändert haben, oder handelt es sich um ein Artefakt? Tatsächlich schränken einige Faktoren die direkte Vergleichbarkeit der beiden Schätzungen ein: (1) Beide Schätzungen sind, wie alle Ergebnisse von Bevölkerungsumfragen, durch Selbstselektion aufgrund von Antwortverweigerung beeinflusst. (2) Während die vorliegende Untersuchung eine Stichprobe der Gesamtbevölkerung nutzen konnte, basierte die damalige Untersuchung auf mehreren regionalen Fallstudien (zudem ausschließlich aus dem Westen Deutschlands), welche jeweils als Zielgebietstichproben konzipiert waren. Die damals gewonnenen Rohdatensätze der Fallstudien wurden mit der jeweiligen Auswahlwahrscheinlichkeit der Besucher gewichtet (vgl. BERNATH *et al.*, 2007); die Hochrechnung der Fallstudien auf die Bundesrepublik erfolgte über eine Metaanalyse („Benefit Transfer“). Es bleibt zu berücksichtigen, dass die damals notwendige Korrektur unterschiedlicher Auswahlwahrscheinlichkeiten eine zusätzliche Fehlerquelle darstellt, und dass auch beim Benefit-Transfer-Ansatz grundsätzlich mit einer höheren Unschärfe der Ergebnisse zu rechnen ist (NAVRUD und READY, 2007; vgl. auch NELSON und KENNEDY, 2009). (3) Dazu kommt, dass die befragten Waldbesucher in beiden Untersuchungen zwar Erfahrung mit dem Bewertungsobjekt „Waldbesuche“ hatten; da die Interviews für die Vorläuferstudien aber im Wald stattfanden, war den Befragten das zu bewertende Objekt im Unterschied zur jetzigen CAPI- und Onlinebefragung unmittelbar präsent.

Alle diese Faktoren können die jeweiligen Bewertungsergebnisse beeinflusst haben. Es ist daher nicht auszuschließen, dass die vorgefundene Differenz im Vergleich zu den früheren Bewertungsergebnissen noch im Bereich der methodischen Messunschärfe liegt. Gleichwohl deuten auch Fallstudienresultate mit einer alternativen Bewertungsmethodik an, dass die derzeitige Bewertung der Erholungsleistung des Waldes niedriger als in den 1990er Jahren liegen könnte.<sup>16)</sup>

<sup>15)</sup> Errechnet über die Inflationsraten der Jahre 1992-2011 (d.s. die Jahre der jeweiligen Datenerhebung), nach Angaben des Statistischen Bundesamtes (<http://de.statista.com/statistik/daten/studie/1046/umfrage/inflationsrate-veraenderung-des-verbraucherpreisindex-zum-vorjahr/>).

<sup>16)</sup> Im Rahmen eines Choice-Experiments (ELSASSER *et al.*, 2010) wurde das Waldbetretensrecht in einem Untersuchungsgebiet im nordostdeutschen Tiefland seitens der dortigen Wohnbevölkerung mit durchschnittlich etwa 75 € pro Haushalt und Jahr bewertet; pro Person umgerechnet sind dies etwa 38 €/p/a. Diese Schätzung übertrifft die in Tabelle 3 für das Bundesgebiet sowie die nordostdeutschen Bundesländer Brandenburg und Mecklenburg-Vorpommern genannten Mittelwerte, die je nach Bezugsbasis zwischen 20 und 32 €/p/a betragen. Die Ergebnisse dieses Choice-Experiments liegen aber deutlich niedriger als selbst die nominalen Werte aus den 1990er Jahren. Diese betragen zwischen 51 und 66 €/p/a je nach Rechenvariante (ELSASSER, 2001:429); inflationsiert entspräche dies gut 70 bis gut 90 €/p/a (s. Fußnote 14).

#### 4.4. Zur zusätzlichen Zahlungsbereitschaft für veränderte Waldbehandlung

Als wesentliche Erweiterung wurde in der aktuellen Studie auch analysiert, welchen Einfluss Veränderungen der Waldbewirtschaftung auf den Erholungswert haben könnten. Diese Fragestellung ist unmittelbarer politikrelevant als nur die Bewertung des Status Quo, weil sie über die möglichen Einflüsse forstpolitischer Gestaltungsmaßnahmen auf den Nutzen der Gesellschaft Auskunft gibt. Die potentiellen Auswirkungen einer (noch) stärker an den Verbesserungswünschen der Bevölkerung orientierten Waldbewirtschaftung auf den Erholungswert erweisen sich aber als gering: Weder eine Ausweitung des infrastrukturellen Angebots (wie sie etwa in den 1970er Jahren betrieben worden ist) noch eine stärkere Naturschutzorientierung (wie heute von vielen Seiten gefordert) würden den ökonomischen Nutzen der Erholung im Wald substantiell erhöhen. Zwar konnte eine Mehrheit der Befragten Verbesserungswünsche benennen, wenn diese unverbindlich abgefragt wurden; selbst hier zeigte sich aber schon, dass ein nicht unerheblicher Anteil gar keine Verbesserungswünsche hat. Fragte man zudem nach den zu erwartenden Auswirkungen entsprechender Verbesserungen auf das (Erholungs-) Verhalten der Befragten, so war es bereits nur noch eine Minderheit, die einschätzte, nach Einführung der von ihnen gewünschten Verbesserungen das Erholungsangebot der Wälder intensiver zu nutzen.<sup>17)</sup> Der Anteil derjenigen schließlich, die hierfür eine zusätzliche Zahlungsbereitschaft hatten, liegt zwischen weniger als zehn und zwanzig Prozent; die Beträge dieser zusätzlichen Zahlungsbereitschaft sind (im Vergleich zum Status Quo) wiederum relativ gering, und zudem müssen die ermittelten Werte als Obergrenzen interpretiert werden, weil die individuellen Verbesserungswünsche, auf die sie sich beziehen, teilweise nicht kompatibel sind.

Damit deutet sich auch an, dass die positiven Nebenwirkungen vieler forstpolitischer Gestaltungsmaßnahmen auf den Erholungswert vernachlässigbar sein könnten – gleich ob es sich dabei um die Förderung des Wegebaus oder die der Naturnähe im Wald handelt. Dies schließt nicht aus, dass in der Bevölkerung substantielle Zahlungsbereitschaften beispielsweise für Naturschutzgüter existieren. In der politischen Diskussion wären naturschutzorientierte Maßnahmen der walddirelevanten Politik aber besser mit ihrem Naturschutznutzen zu begründen, nicht mit ihren Auswirkungen auf den Erholungsnutzen.

#### 5. ZUSAMMENFASSUNG

Zur monetären Bewertung der Erholungsleistung des Waldes in Deutschland wurde 2011 eine Studie durchge-

<sup>17)</sup> Zwar beschränken auch Restriktionen der individuell verfügbaren Freizeit die Nachfrage nach Waldbesuchen. Gleichwohl wäre aus ökonomischer Sicht grundsätzlich zu erwarten, dass vom Befragten gewünschte Angebotsverbesserungen den individuellen Nutzen von Waldbesuchen vergrößern und daher im Rahmen der gegebenen Möglichkeiten zu einer verstärkten Nachfrage führen. Ist dies nicht der Fall, so liegt der Umkehrschluss nahe, dass die gewünschten Angebotsverbesserungen keinen wesentlichen Einfluss auf den Nutzen haben.

führt, deren Bewertungskonzept identisch zu dem mehrerer Vorläuferstudien aus den 1990er Jahren ist. Bewertet wurde mit Hilfe der „Contingent Valuation Method“ jeweils die Zahlungsbereitschaft (ZB) dafür, über ein Jahr hinweg den Wald in der regionalen Umgebung des jeweiligen Wohnortes zur Erholung betreten zu dürfen. Als Erweiterung wurde zusätzlich zum gegebenen Waldzustand auch untersucht, welche Verbesserungen die Bevölkerung wünscht und wie hoch ihre ZB dafür wäre. Als Daten standen 1.011 persönliche Interviews aus einer Bevölkerungsstichprobe sowie zusätzlich 1.049 Antworten einer Internet-Panelbefragung („online“) zur Verfügung.

Gut drei Viertel der Bevölkerung unternehmen mindestens einen Waldbesuch pro Jahr (*Abbildung 2*). Die durchschnittliche ZB der Gesamtbevölkerung für den Status Quo beträgt 26,94 €/p/a, die der Waldbesucher 36,06 €/p/a (*Tabelle 1*); hochgerechnet auf alle Einwohner Deutschlands über 14 Jahren sind dies 1,9 Mrd. €/a (*Tabelle 2*). Mittelwertschätzungen für die 16 Bundesländer schwanken zwischen 11,45 und 42,40 €/p/a (*Tabelle 3*). Die Regressionsanalyse zeigt, dass Unterschiede der ZB wesentlich auf regional unterschiedliche Besuchereigenschaften zurückgehen (v.a. Einkommen, Besuchshäufigkeit); ihre jeweiligen Waldbesuchsaktivitäten haben dagegen keinen durchgängigen Einfluss (*Tabelle 4*).

Etwa die Hälfte der Befragten wünscht sich verschiedene infrastrukturelle Verbesserungen, ein Viertel naturschutzorientierte Verbesserungen; über ein Fünftel hat keine Verbesserungswünsche (*Abbildung 3, Tabelle 5*). Die überwiegende Mehrheit würde aber selbst nach Umsetzung ihrer Verbesserungswünsche keine zusätzlichen Waldbesuche unternehmen; noch weniger äußern eine zusätzliche ZB. Im Durchschnitt aller Befragten beträgt diese zwischen 0,96 und 5,70 €/p/a (*Tabelle 6*). Persönliche und online-Befragung erbringen durchgehend ähnliche, aber nicht gleiche Ergebnisse. Die Größenordnung der ZB stimmen in beiden Befragungsmodi bei allen untersuchten Fragen überein.

Die aggregierte ZB für Walderholung im Status Quo erscheint niedriger als in den 1990er Jahren, was teilweise aber auch methodische Ursachen haben könnte. Hinsichtlich des Verbesserungspotentials lässt sich schließen, dass selbst eine Umsetzung sämtlicher Verbesserungswünsche den Nutzen der Erholung im Wald kaum verbessern würde.

#### 6. ABSTRACT

Title of the paper: *Current and potential recreation value of forests in Germany: monetary benefits of forest recreation from the population's perspective.*

In order to value recreation benefits of the forests in Germany, in 2011 a study was conducted which applied a valuation concept identical to that of several precursor studies from the 1990ies. This application of the Contingent Valuation Method aimed at valuing the willingness to pay (WTP) for the right to access residential and regional forests over one year. Additionally it was asked which improvements are desired by the population, and what their WTP would be for these improvements. Data

were collected by 1.011 personal interviews from a population sample. Additionally 1.049 answers from an internet panel survey were available.

About  $\frac{3}{4}$  of the population visit a forest at least once per year (*figure 2*). The population's mean WTP under the status quo is 26.94 €/p/a; the WTP of forest visitors only is 36.06 €/p/a (*table 1*), which gives an aggregated 1.9 bill. €/a for the whole population of Germany from 14 years of age onwards (*table 2*). Mean estimates for the 16 federal states are between 11.45 und 42.40 €/p/a (*table 3*). The regression analysis reveals that WTP differences are mainly due to regionally different visitor composition (above all, income and visit frequencies). However, their respective recreation activities do not have a dominant influence (*table 4*).

About half of the respondents requested several infrastructural improvements, a quarter requested nature protection oriented improvements; more than one fifth did not have any improvement requests (*figure 3, table 5*). However, the majority would not visit the forests more often than today even if their improvement requests were realised; even fewer people have an additional WTP, which amounts between 0.96 and 5.70 €/p/a (*table 6*). Personal and online-survey consistently yield similar, but not identical results.

The aggregated WTP for forest recreation under the status quo seems lower than in the 1990ies, which might partly be due to methodological causes. With regard to the potential for improvements it can be concluded, that even the realisation of all improvement requests would scarcely increase the aggregated utility of forest recreation.

## 7. RÉSUMÉ

Titre de l'article: *Fonction récréative actuelle et potentielle des forêts en Allemagne: utilité monétaire de la détente en forêt du point de vue de la population.*

Pour permettre l'évaluation de la fonction récréative de la forêt en Allemagne, une étude été menée en 2011. Son concept d'évaluation est identique aux nombreuses études précédentes des années 1990. La disponibilité financière (DF) pour cela a été évaluée à l'aide de la méthode «Contingent Valuation Method», pour avoir le droit de pénétrer pour ses loisirs pendant un an dans la forêt située dans les environs de la région du lieu de résidence. Par extension il a été également recherché, pour un état donné de la forêt, ce que la population souhaite comme amélioration et quel serait le montant de leur disponibilité financière (DF) pour cela. Les données disponibles consistèrent en 1011 interviews personnels issus de sondages dans la population et de manière complémentaire en 1049 réponses à un questionnaire en ligne sur internet.

Environ trois quarts de la population entreprennent au moins une visite forestière par an (*Figure 2*). La disponibilité financière (DF) moyenne de l'ensemble de la population pour le status quo s'élève à 26,94 €/p/a, celle des visiteurs de forêt à 36,06 €/p/a. Ramené à l'ensemble de la population allemande sur 14 ans, ce montant s'élève à 1,9 Milliard €/a (*Tableau 2*). Les estimations des

valeurs moyennes pour les 16 Länder allemands oscillent entre 11,45 et 42,40 €/p/a (*Tableau 3*). L'analyse de régression montre que les différences entre les DF portent essentiellement sur les différences régionales des types de visiteurs (différences de revenu, de fréquence de visite). A chaque fois, les activités des visites en forêt n'ont là-dessus aucune influence dominante (*Tableau 4*).

Environ la moitié des personnes interrogées souhaitent diverses améliorations d'infrastructures, un quart des améliorations orientées dans le sens de la protection de la nature; un peu plus d'un cinquième n'ont aucun souhait d'amélioration (*Figure 3, Tableau 5*). La majorité ne souhaitent cependant pas augmenter leur nombre de visites en forêt en cas de réalisation de leurs souhaits d'améliorations et encore moins le souhaitent avec une augmentation de leur DF. En moyenne, toutes les personnes interrogées estiment leur DF comprise entre 0,96 et 5,70 €/p/a (*Tableau 6*). L'interrogation personnelle et en ligne apporte environ des résultats similaires mais pas identiques. L'ordre de grandeur des DF est le même dans les deux modalités d'interviews.

La DF agrégée pour les loisirs en forêt semble être, dans le status quo, inférieure à ce qu'elle était dans les années 1990. Mais cela pouvait avoir aussi en partie une origine méthodologique. En ce qui concerne le potentiel d'amélioration, on peut conclure que même une réalisation des souhaits d'amélioration améliorerait à peine la fonction «loisirs» de la forêt.

## 8. LITERATUR

- BERGEN, V., W. LÖWENSTEIN und R. OLSCHESKI (2002): Forstökonomie. Volkswirtschaftliche Grundlagen. München: Vahlen. Vahlens Handbücher der Wirtschafts- und Sozialwissenschaften, 469 S.
- BERGEN, V., W. LÖWENSTEIN und G. PFISTER (Hrsg.) (1995): Studien zur monetären Bewertung von externen Effekten der Forst- und Holzwirtschaft. Frankfurt: Sauerländer. Schriften zur Forstökonomie **2**, 185 S.
- BERNATH, K. (2006): Umweltökonomische Bewertung der stadtnahen Walderholung in Zürich. Empirische und Methodische Beiträge zur Analyse von Ziel- und Quellgebietsdaten. Zürich: Universität (Math.-Natwiss. Fak.). Dissertation, 190 S.
- BERNATH, K., P. ELSASSER und A. ROSCHEWITZ (2007): Reduktion systematischer und zufallsbedingter Fehler in Zielgebietsdaten: Theorie und empirische Ergebnisse einer Waldbesucherbefragung in Zürich. Allgemeine Forst- und Jagdzeitung **178** (5/6), S. 109–120.
- BWALDG: Gesetz zur Erhaltung des Waldes und zur Förderung der Forstwirtschaft (Bundeswaldgesetz) vom 2.5.1975 (BWaldG BGBl I, S. 1037), zuletzt geändert 31.7.2010.
- ELSASSER, P. (1996): Der Erholungswert des Waldes. Monetäre Bewertung der Erholungsleistung ausgewählter Wälder in Deutschland. Frankfurt: Sauerländer's. Schriften zur Forstökonomie **11**, 218+25 S.
- ELSASSER, P. (1999): Recreational Benefits of Forests in Germany. In: ROPER, C. S., Park, A. (Hrsg.): The Living Forest. Non- Market Benefits of Forestry. London: The Stationery Office, S. 175–183.
- ELSASSER, P. (2001): Der ökonomische Wert der Wälder in Deutschland für die Naherholung: Eine „Benefit Function Transfer“-Schätzung. Zeitschrift für Umweltpolitik und Umweltrecht **24** (3), S. 417–442.

- ELSASSER, P., H. ENGLERT und J. HAMILTON (2010): Landscape benefits of a forest conversion programme in North East Germany: Results of a choice experiment. *Annals of Forest Research* **53** (1), S. 37–50.
- ELSASSER, P., J. MEYERHOFF, C. MONTAGNÉ und A. STENGER (2009): A bibliography and database on forest benefit valuation studies from Austria, France, Germany, and Switzerland – A possible base for a concerted European approach. *Journal of Forest Economics* **15** (1–2), S. 93–107.
- FARRELL, J. und M. RABIN (1996): Cheap Talk. *Journal of Economic Perspectives* **10** (3), S. 103–118.
- HAGENAARS, A., K. DE VOS und M. A. ZAIDI (1994): *Poverty Statistics in the Late 1980s: Research Based on Microdata*. Luxembourg: Office for Official Publications of the European Communities. Theme 3 Series D, 415 S.
- HALD, A. (1952): *Statistical Theory with engineering Applications*. London: John Wiley & Sons (zit. n. LORENZ, R. (1992): *Grundbegriffe der Biometrie*. Stuttgart/Jena: Fischer. S. 125).
- HSIAO, C. (1983): Regression analysis with a categorized explanatory variable. *In*: KARLIN, S., AMEMIYA, T., GOODMAN, L. (Hrsg.): *Studies in econometrics, time series, and multivariate statistics*. New York: Academic Press, S. 93–129
- KLEIBER, O. (2006): *Monetäre Bewertung von Erholungsnutzen und Nutzerkonflikten in stadtnahen Wäldern. Konzeption und empirische Prüfung am Beispiel des Allschwiler Waldes*. Marburg: Tectum. 296 S.
- KLEIN, C. (1994): *Strategisches Antwortverhalten am Beispiel der Contingent Valuation Method*. Freiburg. Forstwiss. Fak., 53 S.
- KLEIN, C. und P. ELSASSER (1994): *Strategisches Verhalten als mögliche Fehlerquelle der Contingent Valuation Method (CVM)*. *In*: OESTEN, G., ROEDER, A. (Hrsg.): *Zur Wertschätzung der Infrastrukturleistungen des Pfälzerwaldes*. Trippstadt: FVA Rheinland-Pfalz. Mitteilungen der FVA Rheinland-Pfalz **27/94**, S. 111–128.
- LINDHJEM, H. und S. NAVRUD (2011): Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. *International Review of Environmental and Resource Economics* **5** (4), S. 309–351.
- LÖWENSTEIN, W. (1994): *Reisekostenmethode und Bedingte Bewertungsmethode als Instrumente zur monetären Bewertung der Erholungsfunktion des Waldes – Ein ökonomischer und ökonometrischer Vergleich*. Frankfurt: Sauerländer's. Schriften zur Forstökonomie **6**, 206 S.
- LUTTMANN, V. und H. SCHRÖDER (1995): *Monetäre Bewertung der Fernerholung im Naturschutzgebiet Lüneburger Heide*. Frankfurt: Sauerländer's. Schriften zur Forstökonomie **10**, 108 S.
- MARWEDE, M., L. ERDMANN und S. BEHRENDT (2007): *Normative Orientierungen der Wald- und Landnutzung. Leitbild-Assessment im Rahmen des Projektes Zukünfte und Visionen Wald 2100*. Berlin: Institut für Zukunftsstudien und Technologiebewertung. 59 S.
- MITCHELL, R. C. und R. T. CARSON (1990): *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington D.C.: Resources for the Future (2. Aufl.). 463 S.
- NAVRUD, S. und R. READY (Hrsg.) (2007): *Environmental Value Transfer: Issues and Methods*. Springer. The Economics of Non-Market Goods and Resources **9**, 290 S.
- NELSON, J. und P. KENNEDY (2009): *The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment*. *Environmental and Resource Economics* **42** (3), S. 345–377.
- SÖDERQVIST, T. und Å. SOUTUKORVA (2009): On how to assess the quality of environmental valuation studies. *Journal of Forest Economics* **15** (1–2), S. 15–36.
- STBA (Hrsg.) (2011): *Statistisches Jahrbuch für die Bundesrepublik Deutschland*. Wiesbaden: Statistisches Bundesamt, 747 S.

2. Weller, P., und Elsasser, P., 2018. Preferences for forest structural attributes in Germany – Evidence from a choice experiment. *Forest Policy and Economics*, 93, 1-9.



ELSEVIER

Contents lists available at ScienceDirect

# Forest Policy and Economics

journal homepage: [www.elsevier.com/locate/forpol](http://www.elsevier.com/locate/forpol)

## Preferences for forest structural attributes in Germany – Evidence from a choice experiment

Priska Weller\*, Peter Elsasser

Thuenen Institute of International Forestry and Forest Economics, Leuschnerstr. 91, 21031 Hamburg, Germany



### ARTICLE INFO

#### Keywords:

Discrete choice experiment  
Forest  
Willingness to pay  
Mixed logit  
WTP space  
Germany

### ABSTRACT

Using a discrete choice experiment we elicit coefficients and willingness to pay (WTP) values for eight structural characteristics of forests for 2932 respondents from a Germany-wide population sample. Results of the monetary valuation show that positive WTP values exist for increasing the share of forests, increasing biodiversity, increasing the harvest age of forests and maintaining some unused forest areas. Large negative WTP values are found for decreasing the share of forests, changing the amount of coniferous trees to 70%, and increasing the amount of trees from other countries. For some of the attributes investigated changes in any direction are valued negatively, i.e., field size and foreign species. Results are relatively robust to different model specifications, mainly regarding the signs of the preference parameters.

### 1. Introduction

Forests and forested landscapes provide a multitude of goods and services to society. Many of these have characteristics of public goods, for which pure market allocation would fail due to the absence of undistorted market prices. In order to cope with such allocation problems, policies and political strategies have been developed and adopted by governments in Europe (and elsewhere), some of which directly impact on forestry. In Germany, the ‘Forest Strategy 2020’ (BMELV, 2011) aims at accommodating the demands of society towards forests with regard to competing goods and services, such as demands for timber provision, for the mitigation of climate change, for recreation opportunities, and for nature protection. Other sectoral strategies specify particular demands. For example, the German biodiversity and sustainability strategies stipulate that 5% of the forest area be set aside for natural development, that deciduous trees be more favoured than in the past, that the cultivation of non-indigenous tree species be avoided, and that species diversity be better supported (BMU, 2007; RNE, 2002). Several policy instruments for supporting such goals are applied at EU and national level. Among these are legal regulations (e.g., the EU-Habitats Directive, 1992), the Common Agricultural Policy (CAP) and national subsidisation schemes which pay landowners for including some of society’s demands on forests in the enterprises’ management decisions, as well as advisory services of public forest administrations offered to forest landowners. While there are several analyses of the associated costs (Job and Mayer, 2012; Rosenkranz et al., 2014; Wüstemann and Rühls, 2015; Rosenkranz and Seintsch, 2016), studies which account for

the economic benefits of changes in forestal goods and services supply are rare for Germany as a whole, specifically with regard to monetary valuation of public goods associated to biodiversity protection and structural characteristics of forests.

Germany is covered by 11.4 million hectares of forest, summing up to one third of the German area. According to the most recent Federal Forest Inventory (Bundeswaldinventur 3; BMEL, 2015) 48% of forests are privately owned, 19% are corporate forests and 33% are state forests. 54% of trees in German forests are coniferous trees, 46% are deciduous trees. The average age of forests is 77 years, with 24% of forests being > 100 years old (BMEL, 2015). To ensure recreation the general public has the right to enter forests for recreation at any time, including privately owned forests (BWaldG, 2017; §14). German forestry has a long tradition of multifunctionality, meaning that forests are generally used for the production of timber, for the provision of protection services, as well as for recreation purposes at the same time.

Given the widespread multiple demands and trade-offs between different goods and services supplied by forests and society’s demand for these goods and services (see for example BMELV, 2011), economic valuation can support policy making, by identifying options for a more efficient forest resource use that includes society’s preferences towards the various non-marketed public goods. In the policy context shortly described above, this study applies a choice experiment (CE) for valuing public goods provided by forests in Germany at national level. Specifically, we identify the population’s preferences and willingness to pay (WTP) for possible structural changes in eight attributes of forests, which are particularly relevant in current discussions about nature

\* Corresponding author.

E-mail addresses: [priska.weller@thuenen.de](mailto:priska.weller@thuenen.de) (P. Weller), [peter.elsasser@thuenen.de](mailto:peter.elsasser@thuenen.de) (P. Elsasser).

protection in forestry and forest management.

This article is structured as follows: in [Section 2](#) we provide a short literature review, in [Section 3](#) we explain our data base and modelling approach and in [Section 4](#) we present our results. In [Section 5](#) we discuss our results and conclude.

## 2. Literature review

Available valuation studies in Europe deal with valuation of different forest services such as recreation or structure and management of forests, using different elicitation methods (stated preferences like CVM and CE, expert surveys) and in differing (regional) contexts (for a systematic overview of studies from Germany, Austria and Switzerland see [Elsasser et al., 2016a](#); for a general overview over environmental valuation see [Hoyos, 2010](#)). Other studies with a broader regional perspective include [Edwards et al. \(2012\)](#), where European experts assessed preferences for structural attributes of forests for four regions in Europe, or the German TEEB study ([Elsasser et al., 2016b](#); [Bösch et al., 2018](#)), which gives an overview of the economic significance of the goods and services provided by forests, at the same time pointing out uncovered gaps in the valuation of many of these goods and services.

Changing structural characteristics of forests have been the subject of various studies focussing on the value of goods and services like recreation or biodiversity in general. However, an excerpt of what has been elicited in other studies in various contexts about single structural characteristics is explained in the following. As we focus on structural characteristics in forests, we learn from a previous study very similar to ours (a CE with nationwide scope, which has valued agrarian landscape changes with some forest attributes) that respondents stated negative WTP values for decreasing as well as for increasing average sizes of both field and forest patches in the landscape. Furthermore, the mentioned study found that WTP for an increase in the share of forests depends on the abundance of forests in respondents' surroundings, and shows that there are generally large utility gains from an increasing share of forests, but with negative marginal utility after a turning point of the utility function ([Sagebiel et al., 2017](#)). Other European sources also find positive WTP for afforestation, particularly if it is available for recreation, but in a CE in Ireland it was found that location relative to respondents' homes are important for WTP ([Upton et al., 2012](#)). In Italy respondents also have a positive WTP for afforestation, but (as indicated by [Sagebiel et al., 2017](#)) also a mix of landscapes types (particularly 75% woodland, 25% meadows) is valued positively, rather than too much forest ([Vecchiato and Tempesta, 2013](#)).

Some CE studies from different European countries (including Germany) value biodiversity through programmes with attributes like habitats for endangered species, number of species in forests, rare species (of fauna and flora), but also ecosystem processes, forest stand structure or landscape diversity, and find positive WTP values for increases in these attributes ([Christie et al., 2006](#), on farmland in England; [Czajkowski et al., 2009](#), in a nationalpark in Poland; [Meyerhoff et al., 2009](#), in two case study areas in Germany). In a contingent valuation survey the implementation of the national biodiversity strategy in forests is valued at 2.22 billion € ([Meyerhoff et al., 2012](#); [Wüstemann et al., 2014](#)). For ecological reasons, respondents in Poland also have a positive WTP for the protection of forests. This preference proved to be temporally stable and over several choice sets ([Czajkowski et al., 2016](#)). Furthermore, diversity of the general appearance of forests is valued positively: In the UK positive WTP for increases in mixed and broad-leaved forests exist ([Colombo and Hanley, 2008](#)) and in eastern Germany respondents to a CE prefer landscapes with a higher share of forest ([Völker and Lienhoop, 2016](#)). Also, mixed stands are preferred over monocultures and varying tree heights are preferred over uniform tree heights. Overall variation between stands adds to recreation ([Filyushkina et al., 2017](#)). In a non-monetary photo ranking study [Junge et al. \(2011\)](#) found out that the Swiss population prefers a mixed landscape with ecological areas.

Particularly the recreational aspect of forests and its structural attributes is included in several studies. Even results from a recreational context may be interpreted as an indication of preferences about structural attributes. So, forest management has been connected with recreation and investigated by means of CE, finding that more diversity in tree species and tree heights is preferred over monocultures of conifers ([Horne et al., 2005](#); [Nielsen et al., 2007](#); [Elsasser et al., 2010](#); [Dhaka et al., 2012](#); [Abildtrup et al., 2013](#)). Also [Giergiczny et al. \(2015\)](#) investigate a total of 14 management characteristics like forest type, tree species and stand age of forests for recreational purposes in Poland and find in their CE study that more naturalness and diversity in these characteristics increase recreational value, which is measured through distance respondents are willing to travel to recreation forest. Even though more naturalness is desired, forest access restrictions have a negative impact on welfare, although they may be designed to benefit wildlife ([Nielsen et al., 2016](#)). As long as recreation is possible, respondents are willing to pay for the removal of litter and to protect ecologically valuable forests ([Czajkowski et al., 2014](#)).

When we take up what has been done in the various European studies dealing with forest values, we add a large-scale valuation study particularly for entire Germany. We contribute WTP values from a CE for eight structural attributes of forests. The attributes chosen here (see [Section 3.1](#) for details on the attribute selection process) are all closely related to federal German forest policy strategies and programmes. Results from previous studies give us an indication that many of our attributes have been subject to valuation, however in different contexts. We value structural attributes of forests independently from recreation or other services. Specifically, our context is a general landscape change in the surrounding of the respondents' place of living. So our main question is what the German population wants to experience and know to exist in forests and how this is valued monetarily.

## 3. Data and models

### 3.1. Study design and samples

A Germany-wide online survey was conducted in March and April 2013. Respondents were recruited from a representative panel of a survey company,<sup>1</sup> with quotas for age, gender and state of residence. The average interview length was 26 minutes and the response rate<sup>2</sup> was 32.9%. The full questionnaire can be found in the Appendix. However, in the following a short description of the questionnaire is provided.

The main part of the questionnaire consisted of a CE with six different samples, into which each respondent was randomly allocated. In this article, only those two samples are considered which address forest issues. The choice cards within each sample contained six attributes, namely two 'fixed' attributes (Share of forests and Field size) and one price attribute (personal annual contribution to a landscape fund in Euro), which were identical for all samples. Additionally, each choice card included three 'flexible' attributes, which varied between samples. In this way, 'fixed' attributes were valued by 2932 respondents (number of fully completed interviews in Samples 1 and 2), whereas 'flexible' attributes were valued by 1465 (Sample 1) and 1467 (Sample 2) respondents. The attributes and their respective levels are listed in [Table 1](#).

The CE aims at the valuation of eight different forest attributes, which are being intensively discussed in German forest policy. Most attributes refer to characteristics within forests, only the two fixed attributes refer to the mix of land use types. Attributes were chosen based on policy needs and actual discussions about forest utilisation and

<sup>1</sup> LINK Institute, Frankfurt.

<sup>2</sup> The response rate was calculated by dividing the number of interviews completed by the number of persons invited for an interview.

**Table 1**  
Attributes and levels used in the CE.

Names of fixed attributes in survey (in Samples 1 and 2)	Name in models	Levels (status quo: 'as today')
Share of forests	ShareForest_m10	Decrease by 10%
	ShareForest_p10	Increase by 10%
Field and forest size	FieldSize_half	Half the current size
	FieldSize_double	Twice the current size
	Price	0, 10, 25, 50, 80, 110, 160 €
Financial contribution to landscape fund per year		
Flexible attributes in Sample 1	Understorey_half	Half as much as today
	Understorey_double	Twice as much as today
Share of coniferous trees	Conifers_30	30% of forest area
	Conifers_70	70% of forest area
Harvest age of forests	HarvestAge_20	Delay harvest by 20 years
	HarvestAge_30	Delay harvest by 30 years
Flexible attributes in Sample 2		
Biodiversity in forests	Biodiversity_85	Slight increase (to 85 points)
	Biodiversity_105	Considerable increase (to 105 pts.)
Unused forest areas	UnusedAreas_0	0% of forest area
	UnusedAreas_10	10% of forest area
Share of tree species from other countries	ForeignSpecies_half	Half as much as today
	ForeignSpecies_double	Twice as much as today

Note: All attributes had the status quo level 'as today', changes were valued from this base level.

nature protection. Particularly the Forest Strategy 2020 (BMELV, 2011) and the National Biodiversity Strategy (BMU, 2007) were consulted in order to find policy relevant attributes for the CE. The Forest Strategy emphasises, inter alia, the choice of tree species (i.e., reducing the share of coniferous trees and reducing the share of tree species from other countries; leading to attributes 'Conifers' and 'ForeignSpecies'), but also creating unused forest areas with no timber harvest on 5% of the total forest area (attribute 'UnusedAreas'), enhancing biodiversity (attribute 'Biodiversity'), as well as increasing the average age of trees by delaying the harvest age (attribute 'HarvestAge') and enhancing multiple layers forestry (attribute 'Understorey', referring to young trees and shrubs). Furthermore, maintaining or even increasing the share of forest as part of the mix of land uses is one goal of the German Forestry Act (BWaldG, 2017; §1) (fixed attribute 'ShareForest'). The mix of land uses is also addressed in the average size of both field and forest patches in the landscape (fixed attribute 'FieldSize'). The National Biodiversity Strategy (BMU, 2007) additionally defines a target value for the development of biodiversity by means of a biodiversity index (attribute 'Biodiversity'). This biodiversity index measures species diversity and landscape quality by monitoring the population development of characteristic bird species; the sub-indicator for forests embraces eleven bird species (Achtziger et al., 2003). The most recent value available for forests in Germany is on average 87%<sup>3</sup> of the target value (Destatis, 2017). 100 index points are the assumed average biodiversity endowment in the 1970s in Germany and represent the target value for the year 2030.<sup>4</sup> The price attribute was coercive and designed such that each respondent was informed that everyone would have to contribute to a landscape fund, while the exact amount to be payed was still unknown and contingent on the choices being made by the respondents.<sup>5</sup>

<sup>3</sup> At the time the survey was designed, the most recent value for forests was 80 index points (BMU, 2010).

<sup>4</sup> Originally, this was the target value for the year 2015, but the Federal Environment Office has postponed the achievement of the target to the year 2030.

<sup>5</sup> The description of the price attribute was very concise. This may increase likelihood that respondents quickly grasp the information in the text, at the cost of a higher risk for

This payment vehicle was considered easily understandable on the one hand, and more incentive compatible than other commonly understandable alternatives on the other hand (like, e.g., income taxes (Johnston et al., 2017), since a considerable proportion of the German population is not liable to pay income tax). Concerns about misuse of the funds were also addressed and dispelled. Except for the price attribute, all attributes had three levels; these were the status quo level 'as today' and two change levels (In most cases this was one increasing and one decreasing level, predominantly formulated as relative levels referring to the status quo (SQ). However, Conifers, HarvestAge and UnusedAreas had absolute levels, thus being more easily understandable for respondents.). The price attribute had six non-zero levels and one level '0 €' which was presented in the status quo alternative only.

Prior to the CE each attribute was presented with a short description of positive and negative visual effects, though not mentioning political discussions (20–40 words each) and respondents were asked to state their perceived SQ (pSQ) for each attribute on a five point Likert-scale. All attributes presented in the choice cards were illustrated with icons, which facilitated visual discrimination between attributes. Each choice card included two alternatives (Landscape A and Landscape B) with landscape changes plus a status quo alternative (Landscape C), i.e., a zero price alternative with no landscape change. The SQ level 'as today' was included both in the SQ alternative and in the change alternatives on the choice card, the price level 0€ was included only in the SQ alternative. An example choice card is presented in Fig. 1. Respondents were requested to choose their preferred future landscape development on the choice cards considering the landscape within a radius of about 15 km around their place of residence. Each respondent was requested to answer nine choice cards consecutively. Choice cards were presented in a randomised order. After the CE, respondents were asked to state the importance of each attribute for their choice decisions on a five point Likert-scale. At the end of the questionnaire, standard socio-demographic information was requested from respondents.

The qualitative part of the survey, which refers to the pSQ and the importance ratings, will not be presented in this paper due to space limitations. However, in Section 5 they will add to the discussion and interpretation of the WTP results.

Focus group discussions were conducted to test and confirm the relevance and understanding of our attributes; due to the landscape focus of our survey we expected diverging opinions about our attributes and thus we controlled the sampling of focus group participants according to their urban or rural place of residence. We had two focus group discussions with 10 participants each, one in an urban and one in a rural area. Before the CE took place, online pre-tests with 100 participants were conducted, in order to eliminate procedural mistakes.

The reason for the division of respondents and attributes into two samples was to reduce the cognitive burden for the respondents, by reducing the number of attributes to be considered simultaneously. Our decision to split attributes into two different samples rests on the assumption of independence between attributes.

All samples followed a Bayesian c-efficient design which was created using NGENE software. C-efficiency aims at minimizing the sum of the variance of WTP estimates (Scarpa and Rose, 2008). Uniform priors were used, with a positive sign for the environmental attributes and a negative sign for the price attribute.

Generally, descriptive analysis reveals that differences between sample and population means are only moderate. In our survey the mean age of respondents was 43 years (Germany: 50 years for adults) and 47% were female (Germany: 51% for adults). Respondents live in households with an average of 2.6 members and a monthly household net income between 2600€ and 4500€ (about 4/5 of all respondents

(footnote continued)  
possible misunderstandings.

**If only the following options were available for the future development of the landscape within a radius of up to 15 kilometres around your place of residence, which one would you choose?**  
**If you live in a large city, please consider the surrounding area of the city.**

		Landscape A	Landscape B	Landscape C
	<b>Share of forests</b>	Decrease by 10%	Increase by 10%	As today
	<b>Field and forest size</b>	Half the current size	Twice the current size	As today
	<b>Understorey in forests</b>	As today	Twice as much	As today
	<b>Share of coniferous trees</b>	30%	70%	As today
	<b>Harvest age of forests</b>	Delay harvest by 20 years	Delay harvest by 30 years	As today
	<b>Financial contribution to landscape fund per year</b>	25 €	80 €	0 €
<b>I choose</b>				

Fig. 1. Example choice card, translated from German.

specified their respective income; German average for monthly household net income: 2914€ (Destatis, 2015)).<sup>6</sup>

### 3.2. Modelling approach

Model analyses are based on the random utility model (McFadden, 1974; Louviere et al., 2000; Train, 2009), which separates utility ( $U$ ) into a systematic component ( $V$ ) and a random error component for unobservable factors ( $\epsilon$ ) with iid-assumptions. Respondent  $i$  chooses alternative  $j$  from choice set  $n$  which maximises utility and is given by

$$U_{ijn} = V_{ijn} + \epsilon_{ijn} \quad (1)$$

From this we present a multinomial logit model (MNL), which assumes preference homogeneity among respondents. It is a common reference point for choice analyses, and it is considered to be robust to violations of underlying behavioural assumptions (Louviere et al., 2000).

The mixed logit model then takes unobserved heterogeneity into account, which allows for existing taste heterogeneity among respondents. Parameters are assumed to be random, and resulting  $\beta$ s allow for different preferences to be captured by the standard deviation of the random parameters, which allows respondents' preferences to deviate from the mean and thus take variations in results into account. Furthermore, mixed logit models allow the inclusion of the panel character of the data, where one respondent has valued more than one choice card, implying that resulting choices are not necessarily

<sup>6</sup> The inclusion of interactions regarding the influence of income on SQ choices was tested and did not have a significant influence on model results. We consider two possible reasons for this. First, prices may have been too low for preferences to be substantially influenced by budget constraints; second, income variation in our samples might be lower than in the average population (almost 40% of our respondents reported having a university degree, which is generally associated with higher income).

independent of each other (Hensher and Greene, 2003; Colombo et al., 2009; Train, 2009; Bliemer and Rose, 2013).

As WTP is calculated as the ratio of the attribute coefficient and the price coefficient, the WTP from a mixed logit will result in the ratio of two randomly distributed terms. To circumvent the problem of resulting skewed WTP distributions, many researchers leave the price coefficient fixed in the modelling. In order to take into account discussions about the heterogeneity of preferences of respondents regarding prices and the resulting distribution of the price coefficient, some researchers then assume the price coefficient to be random log-normal with only negative parameter values (Hensher and Greene, 2003; Train, 2009). This may also result in highly skewed WTP distributions (Hole and Kolstad, 2012).

In order to avoid this, Train and Weeks (2005) and Daly et al. (2012) recommend using the model in WTP space. With this, assumptions are made for the distribution of WTP directly, not distributions of preferences. Also Scarpa et al. (2008) recommend WTP space models if the distributions of WTP results should be controlled directly.

The separated utility into price  $p$  with a cost parameter  $\alpha$  and a non-price attribute  $x$  with a non-price parameter  $\beta$  can be written as

$$U_{ijn} = -\alpha_i p_{ijn} + \beta_i x_{ijn} + \epsilon_{ijn} \quad (2)$$

The variance of the error term varies over decision makers  $Var(\epsilon_{ijn}) = \lambda_i^2(\pi^2/6)$  and respondent  $i$  has the scale parameter  $\lambda_i$ . Train and Weeks (2005) prove that dividing utility by the scale parameter  $\lambda_i$  does not change behaviour. However, a new error term evolves with the same variance for all respondents. Utility coefficients are now specified as  $\mu_i = \alpha_i/\lambda_i$  and  $c_i = \beta_i/\lambda_i$  which gives a new utility

$$U_{ijn} = -\mu_i p_{ijn} + c_i x_{ijn} + \epsilon_{ijn} \quad (3)$$

Calculating WTP from  $w_i = c_i/\mu_i$  as a ratio between the coefficient of an attribute and the price coefficient gives the WTP space parameterisation

$$U_{ijn} = -\mu_i p_{ijn} + (\mu_i w_i) x_{ijn} + \epsilon_{ijn} \quad (4)$$

This model is scale-free and resulting coefficients of the model can be interpreted straightaway as WTP values which can be compared over different groups or different points in time. In this model specification the price coefficient is assumed to be lognormally distributed (Train and Weeks, 2005; Scarpa et al., 2008; Hole and Kolstad, 2012; Bliemer and Rose, 2013).

Comparisons between models in preference space and WTP space have shown that although models in preference space fit the data better, WTP values tend to be less extreme (Train and Weeks, 2005; Hole and Kolstad, 2012). Scarpa et al. (2008) even found that models in WTP space produce more realistic WTP values and fit the data better.

We use dummy coding for the non-price attributes, where each dummy relates to a reference category (no change/status quo) and an alternative specific constant (ASCsq) which represents the constant of the function. We then use Hole's user-written Stata package for estimating mixed logit model in WTP space with 500 Halton draws in the model estimation (Stata 12, StataCorp).

## 4. Results

### 4.1. Model results

When looking at the choice data, we find that 55% of all choices in Sample 1 (52% in Sample 2) were status quo choices, whereas 45% (48%) of all choices favoured one of the other two alternatives. 419 (351) respondents always opted for the SQ in their choices.<sup>7</sup> Table 2 shows maximum likelihood estimates for the MNL<sup>8</sup> model and for the mixed logit model in WTP space for the two samples. The first column shows results for the MNL model, and for the mixed logit model two columns indicate mean and standard deviation values (SD) for the random coefficients. The upper part of the table shows results for Sample 1, the lower part indicates results for Sample 2.

Generally, the MNL model results in significant coefficients for most non-price attributes. The price coefficient is statistically significant and negative in both samples, indicating a negative influence on choice probability with higher prices. The ASCsq is not significant in both samples, indicating no preference for the status quo. In both samples decreases in ShareForest result in a statistically significant negative coefficient, whereas increases in ShareForest result in a statistically significant positive coefficient. This result is as expected, as the German population visits forests regularly (and passionately).<sup>9</sup> In both samples changes in both directions of FieldSize are valued negatively, which is a noticeable result. Turning to the flexible attributes of Sample 1, we find that a decrease in Understorey has a statistically significant negative coefficient, whereas an increase is not significant. Changing shares of coniferous trees to 30% or 70% leads to statistically significant negative coefficients, indicating that both changes influence the choice probability in a negative direction. Generally, the population assigns a more positive attitude towards deciduous trees. Only the coefficient for an increase in HarvestAge by 20 years is statistically significant and positive, whereas an increase by 30 years is not significant. Turning to the flexible attributes of Sample 2 it can be seen that both levels of increases in Biodiversity are valued statistically significant and positive; furthermore, a large increase in Biodiversity has a larger coefficient than a more modest increase in Biodiversity. This meets our expectations that more Biodiversity results in larger positive preferences. For UnusedAreas only the coefficient for 0% unused areas is statistically significant and negative: Respondents less likely choose alternatives

<sup>7</sup> Those respondents who only chose the status quo were not excluded from our analysis.

<sup>8</sup> These are also published in Fick and Gömann (2018).

<sup>9</sup> From this survey, we know that > 80% of the population spend time outdoors more often than every other day, and for 50% of these, forests are the preferred landscape.

with no unused areas. The coefficients for decreases and increases in the share of ForeignSpecies are both statistically significant and negative, with the coefficient for increases being much larger than that for decreases.

The last two columns show mean and SD results for the mixed logit model for Samples 1 and 2. Mean coefficients can be interpreted straightaway as WTP values, however our model calculation in WTP space has been done with a scaling factor of 1/100 which has to be reversed previous to interpretation (see Section 4.2). Mean coefficients do not diverge substantially from the MNL pattern, however now a change in the share of Conifers to 30% is not statistically significant. Also, in Sample 2 the coefficient for the ASCsq is now statistically significant and negative. Also, changing the share of unused areas to 10% is now statistically significant and positive. Coefficients for Biodiversity show that preferences for a larger increase in Biodiversity are lower than preferences for a smaller increase. Generally, SD values are mostly statistically significant and large, which means that respondents have heterogeneous preferences which influence choices regarding the attributes presented.

Generally, specific values of coefficients cannot be compared across models, due to the different specifications. To facilitate comparisons between models, log likelihood values and information criteria can be found in the last few lines of each part of the table. These indicate that the mixed logit model may be preferred over the basic MNL model, which serves merely as a reference model. The mixed logit model results in lower log likelihood values and better information criteria, which indicates that the inclusion of preference heterogeneity is relevant to improve goodness of fit (Hole and Kolstad, 2012).

### 4.2. WTP values

WTP is calculated for steps from one level (status quo level: as today) to the other respective levels (i.e., for changing the status quo to the situation described by each alternative attribute level). In this section we focus on the mixed logit model, as it results in greater model fit compared to the MNL model.

In Table 3 WTP values for the mixed logit model in WTP space are presented. The upper part of the table shows results from Sample 1, the lower part shows results from Sample 2. In sample 1 there is no statistically significant WTP for the ASCsq, whereas in sample 2 there is a negative WTP for the ASCsq. A negative value indicates preferences to move away from the status quo. Since the status quo varies a lot across Germany, the interpretation of this result is not straightforward. The WTP value for a reduction in ShareForest is statistically significant and negative in both samples, whereas the WTP for an increase in ShareForest is statistically significant and positive in both samples. A decrease in ShareForest generates a larger value than an increase in ShareForest (more than factor 2 and for sample 1 it is the highest absolute value of all attributes and in sample 2 it is still the second highest value). For FieldSize, which is also included in both samples, WTP results for both a decrease and an increase are negative. Here the WTP for a reduction in size is almost twice as large as the WTP result for an increase. WTP estimates do not differ significantly between Samples 1 and 2 (i.e., the mean estimates are inside the respective confidence limits (Pedersen et al., 2014)). This similarity of WTP values between Samples 1 and 2 seems confirming as the respective preferences have been elicited in slightly different contexts. However, it should still be kept in mind that both samples used here treat ShareForest and FieldSize as parts of a forest environment, with only a slightly different focus, created by the varying 'flexible' attributes. In sample 1 the WTP for a reduction in the amount of Understorey is large and negative, whereas an increase in Understorey generates no significant WTP. A change in the share of Conifers<sup>10</sup> to 70% generates a negative WTP,

<sup>10</sup> Cave: Share of coniferous trees has absolute levels: 30% and 70%. The average share

**Table 2**  
Model coefficients from logit models.

	Multinomial logit model		Mixed logit model in WTP space			
			Mean		Standard deviation	
	coef.	z value	coef.	z value	coef.	z value
Sample 1						
ASCsq	0.0717	0.97	-0.1256	-0.86	3.5015***	15.05
ShareForest_m10	-0.4529***	-9.19	-1.2161***	-10.48	1.3573***	8.46
ShareForest_p10	0.3572***	9.47	0.4406***	8.80	0.6537***	7.86
FieldSize_half	-0.2555**	-5.86	-0.6390***	-8.06	0.9040***	8.26
FieldSize_double	-0.2310***	-6.10	-0.3314***	-5.11	0.7546***	8.09
Understorey_half	-0.3872***	-8.17	-0.7453***	-7.74	0.7863***	4.68
Understorey_double	-0.0300	-0.72	0.0262	0.42	-0.9041***	-10.82
Conifers_30	-0.0859*	-1.88	-0.0647	-0.97	0.0066	0.02
Conifers_70	-0.7020***	-15.23	-1.0186***	-8.74	-1.1048***	-11.04
HarvestAge_20	0.2129***	4.18	0.2474***	3.03	0.0225	0.15
HarvestAge_30	0.0226	0.49	-0.0944	-1.28	0.9302***	10.63
Price	-0.0056***	-13.24	0.1419**	2.21	0.1838**	2.14
Log likelihood (Null)	-14,485.2		-			
Log likelihood (Model)	-12,394.36		-8927.501			
Number of observations	39,555		39,555			
Df	12		24			
AIC	24,812.71		17,903.00			
BIC	24,915.74		18,109.05			
Sample 2						
ASCsq	0.0941	1.28	-0.2673*	-1.94	3.3391***	15.66
ShareForest_m10	-0.4354***	-8.49	-1.0949***	-9.41	1.8016***	10.70
ShareForest_p10	0.3115***	8.26	0.4582***	9.33	0.5190***	6.16
FieldSize_half	-0.2550***	-5.84	-0.5617***	-7.28	-0.7601***	-6.34
FieldSize_double	-0.1805***	-4.78	-0.2101***	-3.57	0.6496***	8.02
Biodiversity_85	0.2601***	5.60	0.4464***	6.39	-0.0857	-0.54
Biodiversity_105	0.2967***	7.28	0.3374***	5.54	1.1947***	12.29
UnusedAreas_0	-0.2798***	-6.01	-0.6019***	-7.06	-0.5074***	-2.87
UnusedAreas_10	-0.0225	-0.52	0.1106*	1.91	0.5158***	5.41
ForeignSpecies_half	-0.1590***	-3.38	-0.2222***	-2.89	0.0415	0.18
ForeignSpecies_double	-0.8631***	-18.15	-1.2052***	-9.41	1.0974***	10.76
Price	-0.0047***	-11.46	0.0741	1.20	-0.1851**	-2.53
Log likelihood (Null)	-14,504.98		-			
Log likelihood (Model)	-12,747.9		-9511.488			
Number of observations	39,609		39,609			
Df	12		24			
AIC	25,519.8		19,070.98			
BIC	25,622.84		19,277.06			

\* p < 0.1.  
\*\* p < 0.05.  
\*\*\* p < 0.01.

which is the second highest absolute value in this sample. A change to 30% generates no significant value. Delaying the HarvestAge by 20 years generates some positive WTP, but it is the lowest absolute value in this sample. A delay of HarvestAge by 30 years results in no significant WTP. Turning to the flexible attributes of sample 2 it can be seen that all WTP values are statistically significant. Both WTP values for an increase in Biodiversity are positive, but the WTP for a larger increase is smaller than the WTP for a more moderate increase. This is counterintuitive (furthermore, confidence intervals of the WTP overlap, indicating an almost equal size of WTP values). Changing UnusedAreas to 0% (meaning that all forest areas would be used) generates a large negative WTP, whereas changing UnusedAreas to 10% generates a positive WTP, but it is the lowest value in absolute terms in both samples. WTP values for both a decrease and an increase in ForeignSpecies are negative. The WTP for a decrease is small, but the WTP for an increase is the highest absolute value in sample 2.

(footnote continued)

of conifers is 54% in Germany (BMEL, 2015), with some regional variation. This implies that 30% means a decrease and 70% means an increase for most of the respondents. However, in one federal state (Bundesland Saarland) the actual share of Conifers is lower than 30%, and in another federal state (Bundesland Brandenburg) the actual share of Conifers is higher than 70%.

## 5. Discussion and conclusion

This study elicits preferences and WTP values for eight structural characteristics of forests all over Germany by means of a CE, which was analysed using two different specifications of MNL and mixed logit models. We aim to show results, and relate them to governmental strategies and other studies in this context. Generally, a majority of choices are status quo choices. For some attributes, changes were valued as expected: increases were valued positively and decreases were valued negatively. However, there are specific other attributes for which changes in both directions were valued negatively, namely FieldSize and ForeignSpecies. This indicates that the current status quo endowment was the most favourable status for many respondents with regard to these attributes, and possible changes away from this SQ are refused by a majority. As our attributes were coded as dummies, WTP values should be interpreted as WTP for steps from one attribute level to another.

The WTP values (for confidence limits refer to Table 3) we find show some clear results, and also match the results of the qualitative part of the survey, which are now consulted to improve the interpretation of our WTP results. After the CE, respondents were asked to state subjective levels of importance of attributes for their choices. For an increase in ShareForest we find positive WTP (44€ to 46€ with overlapping confidence intervals), and even higher negative WTP for a

**Table 3**  
Mean WTP values for mixed logit in WTP space.

Sample 1	ll	WTP	ul
ASCsq	–41.26	–12.56	16.14
ShareForest_m10	–144.36	–121.61	–98.87
ShareForest_p10	34.25	<b>44.06</b>	53.87
FieldSize_half	–79.43	–63.90	–48.36
FieldSize_double	–45.86	–33.14	–20.42
Understorey_half	–93.40	–74.53	–55.66
Understorey_double	–9.46	2.62	14.70
Conifers_30	–19.60	–6.47	6.65
Conifers_70	–124.71	–101.86	–79.01
HarvestAge_20	8.72	<b>24.74</b>	40.76
HarvestAge_30	–23.86	–9.44	4.98
Sample 2			
ASCsq	–53.79	–26.73	0.33
ShareForest_m10	–132.29	–109.49	–86.70
ShareForest_p10	36.19	<b>45.82</b>	55.45
FieldSize_half	–71.29	–56.17	–41.05
FieldSize_double	–32.54	–21.01	–9.48
Biodiversity_85	30.94	<b>44.64</b>	58.33
Biodiversity_105	21.79	<b>33.74</b>	45.68
UnusedAreas_0	–76.89	–60.19	–43.49
UnusedAreas_10	–0.30	<b>11.06</b>	22.43
ForeignSpecies_half	–37.28	–22.22	–7.16
ForeignSpecies_double	–145.62	–120.52	–95.42

Note: Statistical significance is indicated in **bold**. ll and ul are lower and upper confidence limits.

decrease (–122€ to –110€ with overlapping confidence intervals), which matches the result that over 90% of respondents found ShareForest ‘very important’ or ‘important’ for their decision. Hence, people are willing to pay for more forest, and definitely do not want to give away the current endowment. Comparing results from samples 1 and 2 for the two fixed attributes shows that for ShareForest WTP results for both levels are very similar with overlapping confidence intervals. For FieldSize an increase is valued higher in Sample 1 than in Sample 2. Furthermore, the ASCsq is significant in Sample 2 only. These differences may be due to small differences in the context generated by the flexible attributes. Both refer to a forest context but with slight differences. Also differences in WTP results may result from sampling effects. Also the WTP values for Conifers\_70 (–102€) and ForeignSpecies\_double (–121€) are among the largest absolute values, but importance ratings are ambiguous. Particularly for Conifers the largest group of respondents (37%) stated that it was ‘neither important nor unimportant’ for their decision. ForeignSpecies was also rated rather unimportant by respondents (only 23% stated it was ‘very important’ or ‘important’, 47% stated it was ‘unimportant’ or ‘very unimportant’). The very high negative WTP is surprising and may be generated by a small group of respondents with strong negative preferences, whereas the majority of respondents did not assign much importance to this attribute. Interesting (from a policy perspective) is that reducing UnusedAreas to zero generates large negative WTP values (–60€), whereas a change to 10% of forest areas (which is an increase on average) is valued only with low WTP values (11€). This can be interpreted such that using all forest areas is not preferred, but a change to 10% is not very important to respondents. For this attribute > 50% of respondents stated that they found this ‘very important’ or ‘important’ for their decision. For Biodiversity even 85% stated that they found this attribute ‘very important’ or ‘important’ for their decision, however, resulting WTP is positive, but not as high as this result suggests (34€ to 45€ with overlapping confidence intervals). A reduction in Understorey is clearly valued negatively (–75€), although around 30% of respondents stated that it was ‘neither important nor unimportant’ for their decision. For HarvestAge also around 30% of respondents stated that it was ‘neither important nor unimportant’ for their decision, and resulting WTP values are low for a 20 year increase (25€) and insignificant for a 30 year increase in HarvestAge. For FieldSize changes

in both directions are valued negatively in both samples (–63€ to –56€ with overlapping confidence intervals for a decrease and –33€ to –21€ for an increase), but > 50% found this ‘very important’ or ‘important’ for their decision. Generally, heterogeneity in initial endowment (and consecutively, heterogeneity in results) is large, meaning that local preferences can be much more distinct, even if the mean values reported here suggest differently.

One further explanation of what generated some of the results may be that for some attributes, the assumption of a monotonous relationship between attribute changes and utility may be challenged. In general economic contexts monotony is a straightforward and reasonable assumption. However, in a landscape valuation context this might ignore preferences against changing the status quo in any direction. This may be the case when finding results like those for FieldSize and ForeignSpecies, where changes in both directions (decrease and increase) are valued negatively.

Generally, CE results are mean values over all respondents located all over Germany, and inclusion of pronounced regional variability of forest cover and composition may change results (Domínguez-Torreiro and Soliño, 2011; Czajkowski et al., 2015). Unfortunately, for many attributes objective SQ data are not available for Germany, which would be needed for such an analysis. Where data are available, for example Sagebiel et al. (2017) have included these in a regionally explicit WTP model and found dependency on attribute endowment: WTP for afforestation is lower where the initial endowment with forests is higher. In the qualitative part of our survey respondents were also asked to state their perceived status quo levels of attributes, but mostly the category ‘medium’ was chosen, which we assume indicates a lack of knowledge.

With regard to the Forest strategy and the National Biodiversity strategy of the German Federal Government we see that the results of the population survey do not contradict the aims stated there on behalf of the population. This means that aiming for maintaining or even increasing the amount of forest in Germany, more biodiversity, maintaining a substantial amount of deciduous trees in the forest area, keeping some forest areas unused, a lower share of trees from other countries and more multiple layers forestry meets the preferences of the German population. When comparing our results to other papers which have investigated and valued structural attributes of forests by means of CE, we see that our results are similar to others. Particularly our negative WTP results for FieldSize and positive WTP values for increases in ShareForest are confirmed by others (Sagebiel et al., 2017). Besides this, a nationwide population sample in adjacent Poland expressed preferences for higher forest stand ages (70 and 100 years, with the former being already the average age of forest stands in Germany), where in Germany only increasing the age of forests by 20 years generated positive WTP values. Furthermore, preferences for stands with four broadleaved tree species detected in Poland (Giergiczny et al., 2015) correspond with our results, where increases in conifers are also valued negatively. Also, results from the Swiss lowlands (Junge et al., 2011) match our finding of a preference against a large share of coniferous trees and for more biodiversity.

When comparing the reference model and the mixed logit model in WTP space, we see that the model allowing for preference heterogeneity dominates the MNL model with assumed preference homogeneity in terms of model fit. Signs of preferences however, mostly show in the same direction. Choice of mixed logit model in WTP space was made due to the fact that WTP space models often produce less extreme WTP results, which may be advantageous in policy context (Train and Weeks, 2005). When basing policy advice on valuation results this may be an attractive feature. Results of CE are subject to many factors and there is not ‘one’ WTP value but a range of elicited values, which originate from the method of CE. Regarding the future use of results, all aggregations of values for political use have to be done with caution (Bateman et al., 2006).

This study adds values for eight structural forest attributes to the

existing literature, elicited all over Germany. For many of these attributes there are no values from previous studies in Germany. These attributes are closely linked to several political strategies for landscape development, which will shape land use in the country in future years. So, preferences of the German population can be taken into account when deducing measures and actions from political strategies. Hence, our results provide a valuable basis for political discussions about landscape changes.

## Acknowledgements

The authors wish to thank two anonymous reviewers for very valuable comments. The authors also wish to thank Jürgen Meyerhoff for helpful discussions and comments. Funding for this research was provided by the German Federal Ministry of Education and Research (Fkz. 01LL0909A, project CC-LandStrad) and the German Federal Ministry of Food and Agriculture (Fkz. FNR-22022614, project ReWaLe) and is gratefully acknowledged.

## Appendix A. Supplementary data

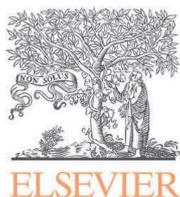
Supplementary data to this article can be found online at <https://doi.org/10.1016/j.forpol.2018.04.013>.

## References

- Abildtrup, J., Garcia, S., Olsen, S.B., Stenger, A., 2013. Spatial preference heterogeneity in forest recreation. *Ecol. Econ.* 92, 67–77.
- Achtziger, R., Stickroth, H., Zieschank, R., 2003. F + E-Projekt "Nachhaltigkeitsindikator für den Naturschutzbereich". *Berichte des Landesamtes für Umweltschutz Sachsen-Anhalt. Sonderheft 1* (2003), 138–142.
- Bateman, I.J., Day, B.H., Georgiou, S., Lake, I., 2006. The aggregation of environmental benefit values: welfare measures, distance decay and total WTP. *Ecol. Econ.* 60 (2), 450–460.
- Bliemer, M.C., Rose, J.M., 2013. Confidence intervals of willingness-to-pay for random coefficient logit models. *Transp. Res. B Methodol.* 58, 199–214.
- BMEL, 2015. *Bundeswaldinventur 3*. [www.bundeswaldinventur.de](http://www.bundeswaldinventur.de) (Berlin).
- BMELV, 2011. *Forest Strategy 2020. Sustainable Forest Management – An Opportunity and a Challenge for Society*. BMELV, Bonn.
- BMU, 2007. *National Strategy on Biological Diversity*. Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Bonn.
- BMU, 2010. *Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt*. Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Bonn.
- Bösch, M., Elsasser, P., Franz, K., Lorenz, M., Moning, C., Olschewski, R., Rödl, A., Schneider, H., Schöppel, B., Weller, P., 2018. Forest ecosystem services in rural areas of Germany: insights from the national TEEB study. *Ecosyst. Serv.* 31, 77–83.
- BWaldG, 2017. (Gesetz zur Erhaltung des Waldes und zur Förderung der Forstwirtschaft, Bundeswaldgesetz) vom 2. Mai 1975 (BGBl. I S. 1037), zuletzt geändert durch Artikel 1 des Gesetzes vom 17. Januar. (BGBl. I S. 75). 13 S.
- Christie, M., Hanley, N., Warren, J., Murphy, K., Wright, R., Hyde, T., 2006. Valuing the diversity of biodiversity. *Ecol. Econ.* 58 (2), 304–317.
- Colombo, S., Hanley, N., 2008. How can we reduce the errors from benefits transfer? An investigation using the choice experiment method. *Land Econ.* 84 (1), 128–147.
- Colombo, S., Hanley, N., Louviere, J., 2009. Modeling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture. *Agric. Econ.* 40 (3), 307–322.
- Czajkowski, M., Buszko-Briggs, M., Hanley, N., 2009. Valuing changes in forest biodiversity. *Ecol. Econ.* 68 (12), 2910–2917.
- Czajkowski, M., Bartzak, A., Giergiczy, M., Navrud, S., Żylicz, T., 2014. Providing preference-based support for forest ecosystem service management. *Forest Policy Econ.* 39, 1–12.
- Czajkowski, M., Budziński, W., Campbell, D., Giergiczy, M., Hanley, N., 2015. Spatial Heterogeneity of Willingness to Pay for Forest Management (No. 2015-07).
- Czajkowski, M., Barczak, A., Budziński, W., Giergiczy, M., Hanley, N., 2016. Preference and WTP stability for public forest management. *Forest Policy Econ.* 71, 11–22.
- Daly, A., Hess, S., Train, K., 2012. Assuring finite moments for willingness to pay in random coefficient models. *Transportation* 39 (1), 19–31.
- Destatis, 2015. *Wirtschaftsrechnungen - Einkommens- und Verbrauchsstichprobe - Einnahmen und Ausgaben privater Haushalte, Fachserie 15, Heft 4, Artikelnummer: 2152604139004*. Statistisches Bundesamt, Wiesbaden.
- Destatis, 2017. *Nachhaltige Entwicklung in Deutschland, Indikatorenbericht 2016*. Bestellnummer: 0230001-16900-1. Statistisches Bundesamt, Wiesbaden.
- Dhakil, B., Yao, R.T., Turner, J.A., Barnard, T., 2012. Recreational users' willingness to pay and preferences for changes in planted forest features. *Forest Policy Econ.* 17, 34–44.
- Domínguez-Torreiro, M., Soliño, M., 2011. Provided and perceived status quo in choice experiments: implications for valuing the outputs of multifunctional rural areas. *Ecol. Econ.* 70 (12), 2523–2531.
- Edwards, D., Jay, M., Jensen, F.S., Lucas, B., Marzano, M., Montagné, C., Weiss, G., 2012. Public preferences for structural attributes of forests: towards a pan-European perspective. *Forest Policy Econ.* 19, 12–19.
- Elsasser, P., Englert, H., Hamilton, J., 2010. Landscape benefits of a forest conversion programme in north East Germany: results of a choice experiment. *Ann. For. Res.* 53 (1), 3.
- Elsasser, P., Meyerhoff, J., Weller, P., 2016a. *An Updated Bibliography and Database on Forest Ecosystem Service Valuation Studies in Austria, Germany and Switzerland*. Johann Heinrich von Thünen-Institut, Braunschweig (20 p, Thünen Working Paper 65). <https://doi.org/10.3220/WP1479222082000>.
- Elsasser, P., Kawaletz, H., Bormann, K., Bösch, M., Lorenz, M., Moning, C., Olschewski, R., Rödl, A., Schröppel, B., Weller, P., 2016b. Ökosystemleistungen von Wäldern. In: *Ökosystemleistungen in ländlichen Räumen: Grundlage für menschliches Wohlergehen und nachhaltige wirtschaftliche Entwicklung; Naturkapital Deutschland - TEEB.de*. BfN, Bonn, pp. 4–31.
- EU-Habitats Directive, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. In: *Official Journal of the European Union* L 206, 22/07/1992, 0007-0050.
- Fick, J., Gömann, H. (Eds.), 2018. *Wechselwirkungen zwischen Landnutzung und Klimawandel*. Springer, Heidelberg (in preparation).
- Filyushkina, A., Agimass, F., Lundhede, T., Strange, N., Jacobsen, J.B., 2017. Preferences for variation in forest characteristics: does diversity between stands matter? *Ecol. Econ.* 140, 22–29.
- Giergiczy, M., Czajkowski, M., Żylicz, T., Angelstam, P., 2015. Choice experiment assessment of public preferences for forest structural attributes. *Ecol. Econ.* 119, 8–23.
- Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. *Transportation* 30 (2), 133–176.
- Hole, A.R., Kolstad, J.R., 2012. Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empir. Econ.* 42 (2), 445–469.
- Horne, P., Boxall, P.C., Adamowicz, W.L., 2005. Multiple-use management of forest recreation sites: a spatially explicit choice experiment. *For. Ecol. Manag.* 207 (1), 189–199.
- Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. *Ecol. Econ.* 69 (8), 1595–1603.
- Job, H., Mayer, M., 2012. Forstwirtschaft versus Waldnaturschutz: Regionalwirtschaftliche Opportunitätskosten des Nationalparks Bayerischer Wald. *AfJZ* 183, 129–143.
- Johnston, R.J., et al., 2017. Contemporary guidance for stated preference studies. *J. Assoc. Environ. Resour. Econ.* 4 (2), 319–405.
- Junge, X., Lindemann-Matthies, P., Hunziker, M., Schüpbach, B., 2011. Aesthetic preferences of non-farmers and farmers for different land-use types and proportions of ecological compensation areas in the Swiss lowlands. *Biol. Conserv.* 144 (5), 1430–1440.
- Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York, NY, pp. 105–142.
- Meyerhoff, J., Liebe, U., Hartje, V., 2009. Benefits of biodiversity enhancement of nature-oriented silviculture: evidence from two choice experiments in Germany. *J. For. Econ.* 15 (1), 37–58.
- Meyerhoff, J., Angeli, D., Hartje, V., 2012. Valuing the benefits of implementing a national strategy on biological diversity — the case of Germany. *Environ. Sci. Pol.* 23, 109–119.
- Nielsen, A.B., Olsen, S.B., Lundhede, T., 2007. An economic valuation of the recreational benefits associated with nature-based forest management practices. *Landsc. Urban Plan.* 80 (1), 63–71.
- Nielsen, A.S.E., Lundhede, T.H., Jacobsen, J.B., 2016. Local consequences of national policies - a spatial analysis of preferences for forest access reduction. *Forest Policy Econ.* 73, 68–77.
- Pedersen, L.B., Riise, J., Hole, A.R., Gyrd-Hansen, D., 2014. GPs' shifting agencies in choice of treatment. *Appl. Econ.* 46 (7), 750–761.
- RNE, 2002. *Perspectives for Germany. Our Strategy for Sustainable Development*. RNE, Berlin.
- Rosenkranz, L., Seintsch, B., 2016. Opportunitätskostenanalyse zur Implementierung des naturschutzorientierten Waldbehandlungskonzepts "Neue Multifunktionalität". *Landbauforschung* 65, 145–160.
- Rosenkranz, L., Seintsch, B., Wippel, B., Dieter, M., 2014. Income losses due to the implementation of the habitats directive in forests — conclusions from a case study in Germany. *Forest Policy Econ.* 38, 207–218.
- Sagebiel, J., Glenk, K., Meyerhoff, J., 2017. Spatially explicit demand for afforestation. *Forest Policy Econ.* 78, 190–199.
- Scarpa, R., Rose, J.M., 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Aust. J. Agric. Resour. Econ.* 52 (3), 253–282.
- Scarpa, R., Thieme, M., Train, K., 2008. Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the alps. *Am. J. Agric. Econ.* 90 (4), 994–1010.
- Train, K.E., 2009. *Discrete Choice Methods with Simulation*, 2nd edition. Cambridge University Press.
- Train, K., Weeks, M., 2005. Discrete choice models in preference space and willingness-to-pay space. In: *Applications of Simulation Methods in Environmental and Resource Economics*, pp. 1–16.
- Upton, V., Dhubbain, A.N., Bullock, C., 2012. Preferences and values for afforestation: the effects of location and respondent understanding on forest attributes in a labelled

- choice experiment. *Forest Policy Econ.* 23, 17–27.
- Vecchiato, D., Tempesta, T., 2013. Valuing the benefits of an afforestation project in a peri-urban area with choice experiments. *Forest Policy Econ.* 26, 111–120.
- Völker, M., Lienhoop, N., 2016. Exploring group dynamics in deliberative choice experiments. *Ecol. Econ.* 123, 57–67.
- Wüstemann, H., Rühls, M., 2015. Was kostet der Naturschutz in Deutschland? In: Eine Spezifizierung des Finanzbedarfs, aktueller Ausgaben und Finanzierungslücken. *Zeitschrift für Umweltpolitik und Umweltrecht.* vol. 38, pp. 27–52.
- Wüstemann, H., Meyerhoff, J., Rühls, M., Schäfer, A., Hartje, V., 2014. Financial costs and benefits of a program of measures to implement a National Strategy on biological diversity in Germany. *Land Use Policy* 36, 307–318.

3. Weller, P., Oehlmann, M., Mariel, P., und Meyerhoff, J., 2014. Stated and inferred attribute non-attendance in a design of designs approach. *Journal of choice modelling*, 11, 43-56.



Contents lists available at ScienceDirect

# The Journal of Choice Modelling

journal homepage: [www.elsevier.com/locate/jocm](http://www.elsevier.com/locate/jocm)



## Stated and inferred attribute non-attendance in a design of designs approach



Priska Weller<sup>a,\*</sup>, Malte Oehlmann<sup>b</sup>, Petr Mariel<sup>c</sup>, Jürgen Meyerhoff<sup>b</sup>

<sup>a</sup> Johann Heinrich von Thünen Institute, Institute for International Forestry and Forest Economics, Leuschnerstraße 91, 21031 Hamburg, Germany

<sup>b</sup> Technische Universität Berlin, Institute for Landscape and Environmental Planning, Straße des 17. Juni 145, 10623 Berlin, Germany

<sup>c</sup> University of the Basque Country (UPV/EHU), Department of Applied Economics III (Econometrics and Statistics), Avda. Lehendakari Aguirre, 83, 48015 Bilbao, Spain

### ARTICLE INFO

#### Article history:

Received 14 December 2013

Received in revised form

12 April 2014

Accepted 25 April 2014

Available online 16 June 2014

#### Keywords:

Attribute non-attendance

Design of designs

Stated choice experiment

Latent class analysis

### ABSTRACT

Attribute non-attendance in stated choice experiments (CE) has gained attention in literature, with some studies finding that not all respondents attend to all attributes. While the current studies show that taking non-attendance into account can significantly influence survey results, it is not yet clear what motivates respondents to ignore or pay less attention to some of the attributes. In the present study, we use 16 different split samples designed according to a design of designs plan, varying different aspects of dimensionality, i.e., the number of choice sets, the number of alternatives, or the number of attributes. Firstly, to analyse the relationship between stated attribute non-attendance and the design dimensions we test whether both are significantly associated. Secondly, we estimate equality-constrained latent class models with classes based on pre-defined rules to infer attribute non-attendance and analyse the influence of the design dimensions. Overall, the results indicate a rather weak relationship between stated or inferred attribute non-attendance and design dimensions. However, an interesting finding is that there is an association with the number of alternatives and with the number of sets.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Recently, the question whether respondents to stated choice experiments (CE) attend to all attributes presented on the choice sets has gained quite a bit of interest. Results from several studies indicate that not all respondents attend to all attributes and that attribute non-attendance (ANA) can significantly bias parameter and subsequent willingness to pay (WTP) estimates (Hensher et al., 2005; Scarpa et al., 2009; Campbell et al., 2011). The awareness that ANA can be an issue when estimating discrete choice models leads to the question of the causes of ANA. One source of ANA might be the complexity of the CE as respondents may employ different answer heuristics when the choice sets become more complex (Scarpa et al., 2009; Hensher et al., 2012). So, respondents may focus only on a subset of attributes.

In the present study we interpret the complexity of the choice task as the design dimensionality of the CE. We systematically vary the number of choice sets, the number of alternatives, the number of attributes and their levels as well

\* Corresponding author.

E-mail addresses: [priska.weller@ti.bund.de](mailto:priska.weller@ti.bund.de) (P. Weller), [malte.oehlmann@tu-berlin.de](mailto:malte.oehlmann@tu-berlin.de) (M. Oehlmann), [Petr.Mariel@ehu.es](mailto:Petr.Mariel@ehu.es) (P. Mariel), [juergen.meyerhoff@tu-berlin.de](mailto:juergen.meyerhoff@tu-berlin.de) (J. Meyerhoff).

as the level range following a Design of Designs (DoD) approach introduced by [Hensher \(2004\)](#) in transportation research. Using 16 split samples, this study aims to investigate the relationship between these five design dimensions and the attendance to attributes. Influences of choice task complexity have been analysed in terms of several issues such as error variance (e.g., [Caussade et al., 2005](#); [Meyerhoff et al., 2014](#)) or the number of status quo choices (e.g., [Boxall et al., 2009](#)). However, we are not aware of any study systematically relating ANA to the design dimensionality of the CE in random utility models (RUMs). To our knowledge only [Hensher \(2006\)](#) investigated choice task complexity influences on ANA analysing stated non-attendance data. Being aware of the design dimensions that cause ANA might be important for the researcher not only in the estimation stage, but particularly in the design stage of the CE.

The data used in this study comes from a nation-wide online survey on land use changes in Germany. The attributes we considered were: Share of Forest, Land Conversion and different biodiversity attributes which were aggregated and disaggregated across designs according to different landscape types. This allowed us to systematically focus on the effect of the design dimensions. After answering the choice questions we asked respondents to state their attendance to the attributes they were presented with in the CE. In total we collected 1684 interviews resulting in an average of more than 100 respondents per treatment.

We analyse the relationship between ANA and the design dimensionality in two different ways. Firstly, we report the answers to non-attendance questions across the 16 different designs and identify those design dimensions that significantly influence stated ANA. Secondly, we model inferred ANA by specifying an equality-constrained latent class model (ECLCM) for each design. We then derive inferences of dimensionality impacts on ANA by comparing the membership probabilities of different ANA classes across the 16 LC models. We apply this two-step approach separately analysing stated and inferred ANA since it has been argued that even when respondents stated that they have not attended to an attribute, the models indicate that people have at least partially taken them into account (e.g., [Hensher and Rose, 2009](#)). Furthermore, both [Scarpa et al. \(2013\)](#) and [Kragt \(2013\)](#) compared two approaches, one not incorporating and one incorporating stated ANA. [Kragt \(2013\)](#) did this by setting attribute parameters to zero where ANA had been stated by respondents. They found that the constrained LC model, which did not include stated ANA, better matched observed data.

The paper is structured as follows: a literature review is presented in [Section 2](#). [Section 3](#) then explains the design of the study and the sampling procedure. Next, [Section 4](#) outlines the hypotheses to be tested in this study and the modelling approach while [Section 5](#) presents results. [Section 6](#) concludes with a discussion of the results and their implications.

## 2. Literature review

The literature on attribute non-attendance in stated choice studies can be divided broadly into two main approaches: the stated non-attendance approach and the analytical non-attendance approach. In the Appendix Table A1 with studies investigating ANA is provided. In the former, respondents are asked directly whether they have considered all attributes describing the alternatives of the choice tasks or whether they have ignored one or more attributes while choosing among them. Stated ANA is usually investigated using a binary response variable with the information on ANA commonly being collected after answering all choice sets (serial stated ANA). Alternatively, non-attendance questions may be asked after each choice set (e.g., [Scarpa et al., 2009](#)). The answers to these questions are then used to put certain restrictions on RUMs. Among the first studies investigating ANA was [Hensher et al. \(2005\)](#). They used the answers to a binary non-attendance question to specify a mixed logit model in which respondents who stated that they ignored a certain attribute were expected to have zero utility. This approach of individual-level zero marginal utility weights has subsequently been applied in several other studies including [Hensher \(2006\)](#), [Hensher et al. \(2007\)](#) and [Kragt \(2013\)](#). Another way to implement the information gained from a stated non-attendance question into RUMs is to estimate different coefficients for each attribute: one for the group of respondents who stated that they did not ignore the attributes and one for those who stated that they had ([Hess and Hensher, 2010](#); [Scarpa et al., 2013](#); [Colombo et al., 2013](#)).

However, the reliability of the stated ANA approach has been put into question. Firstly, it has been argued that a respondent may assign low importance to an attribute although stating that he or she ignored it completely (e.g., [Hess and Hensher, 2010](#)). This would lead to an overestimation of ANA ([Carlsson et al., 2010](#)). Nevertheless, it should be noted that in some instances it was found that respondents who report having ignored an attribute do indeed have zero marginal utility for that attribute ([Balcombe et al., 2011](#)). Secondly, directly incorporating the responses to the non-attendance questions into the RUM may cause potential problems of endogeneity bias ([Hess and Hensher, 2010](#)). Thirdly, there is also an increase in literature raising the – still unanswered – question of the reliability of complementary questions regarding non-attendance ([Hensher and Rose, 2009](#); [Hensher and Greene, 2010](#)).

Given the limitations of the stated non-attendance approach, there has been increasing interest in literature which analyses different methods of inferring attribute non-attendance. These studies infer the information from the data by using diverse econometric treatments. The modelling approach that has probably been implemented most consists of LC models in which a probabilistic decision process captures the attendance to attributes imposing specific restrictions on the utility expressions for each class. The majority of studies using discrete probability distributions implemented the ECLCM, where ANA is operationalised by allowing some respondents to belong to latent classes with zero utility weights for selected attributes, while non-zero parameters are assumed to take the same values across classes ([Scarpa et al., 2009, 2013](#); [Hensher and Greene, 2010](#); [Campbell et al., 2011](#); [Hess et al., 2013](#)).

Nevertheless, [Hess et al. \(2013\)](#) argue that the latent class approach so far applied might be misguided as the results might be confounded with regular taste heterogeneity. They suggest instead a combined LC mixed logit model that allows jointly for attribute non-attendance and continuous taste heterogeneity. One of their findings is that non-attendance is substantially reduced in these models. Similarly, [Hensher et al. \(2013\)](#) presented different latent class models accounting not only for ANA, but also for aggregation of common-metric attributes. One of their models comprises several full attendance classes while other classes account for ANA regarding certain attributes. The model does not constrain the parameters to be equal across classes. In the next step they add to this model an additional layer of heterogeneity by specifying some parameters to follow a random distribution. For their data they found that accounting additionally for taste heterogeneity through random parameters within a latent class only marginally improves the model, but increases the probability of membership to full attribute attendance classes. [Collins et al. \(2013\)](#) presented a generalised random parameters ANA model (RPANA) as an alternative approach; besides, they found that with stated ANA as covariates the model performance was improved.

Studies that employ both the stated non-attendance and the inferred non-attendance approach are [Hensher et al. \(2007\)](#), [Hensher and Rose \(2009\)](#), [Campbell et al. \(2011\)](#), [Kragt \(2013\)](#) and [Scarpa et al. \(2013\)](#). The overall finding is that results from analytical and stated ANA are not consistent, and that the inferred approach provides a better model fit. However, independent of the modelling approach almost all studies find that accounting for ANA improves model fit, indicating that respondents have indeed applied different processing strategies ([Hensher et al., 2005](#); [Hess and Hensher, 2010](#); [Scarpa et al., 2013](#)). Moreover, accounting for ANA leads to significantly different willingness to pay estimates which may be lower ([Hensher, 2006](#); [Campbell et al., 2008](#); [Scarpa et al., 2009](#)) or higher ([Hensher et al., 2005](#); [Hensher and Rose, 2009](#); [Hensher and Greene, 2010](#); [Lagarde, 2013](#)).

Systematic relationships between ANA and the design dimensionality of the CE have so far only been investigated in one study. [Hensher \(2006\)](#) used an ordered heterogeneous logit model with the dependent variable defining the number of attributes that were stated to be ignored, and the independent variables being, among others, the dimensionality. [Hensher \(2006\)](#) found evidence that all design dimensions influence stated ANA. However, we are not aware of any study that relates ANA to the design dimensionality of the CE by either implementing the responses to a stated ANA question into a RUM or inferring ANA from the data.

### 3. Study design and samples

The non-price attributes used in the stated CE are all related to environmental aspects associated with land use changes. The list of attributes ([Table 1](#)) comprises Share of Forest, Land Conversion, several attributes regarding biodiversity conservation and a price attribute Cost. Cost is presented as an annual contribution to a landscape fund. All attributes except those concerning biodiversity conservation were presented in all designs. Moreover, for Cost attribute levels were not varied across designs. To adjust the number of attributes according to the design plan, the attribute Biodiversity Conservation was based on an indicator using stocks of bird populations in Germany ([BMU, 2010](#)). This indicator can be segregated and aggregated into various bird populations in different landscapes (e.g., birds in the whole landscape can be split up into birds in agrarian landscape plus birds in other landscapes).

Following the design master plan by [Hensher \(2004\)](#), 16 different designs were created using NGENE software. The numbers of choice sets used were 6, 12, 18 or 24, and we presented three to five alternatives including a status quo option. The number of attributes ranged from four to seven. The number of attribute levels varied from two to four and the range in attribute levels had three specifications: A base range, a narrow range (base  $-20\%$ ), and a wide level range (base  $+20\%$ ). From these dimensions, C-efficient designs were created. They minimise the variance of the marginal WTP estimates ([Scarpa and Rose, 2008](#)). Uniform priors were used, ranging from 0.1 to 1.2 with a positive sign for the non-cost attributes and a negative sign for Cost. However, due to time constraints we did not update the design while the survey was conducted. Each respondent was randomly allocated to one of the designs. In [Table 2](#) the different combinations of design dimensions are presented; the design numbers (1–16) indicate the sequence as seen in [Hensher \(2004\)](#).

**Table 1**  
Attributes used in the CE.

Attribute	Description
Share of Forest	Percentage changes in the Share of Forest
Land Conversion	Percentage changes in land conversion for housing development and traffic
Bio_whole	Biodiversity in the whole landscape including all landscape types
Bio_agrar	Agricultural landscape biodiversity
Bio_forest	Forest landscape biodiversity
Bio_urban	Urban area biodiversity
Bio_other1	Biodiversity in other landscape types: Forests, urban areas, mountains, waters
Bio_other2	Biodiversity in other landscape types: Urban areas, mountains, waters
Bio_other3	Biodiversity in other landscape types: Mountains, waters
Cost	Contribution to a landscape fund in € per year

**Table 2**  
Design overview.

Design	Sets	Alternatives	Attributes	Levels	Range
1	24	4	5	3	Base
2	18	4	5	4	Base + 20%
3	24	3	6	2	Base + 20%
4	12	3	6	4	Base
5	6	3	4	3	Base + 20%
6	24	3	4	4	Base – 20%
7	6	4	7	2	Base – 20%
8	12	5	4	4	Base + 20%
9	24	5	4	4	Base
10	6	5	7	3	Base + 20%
11	6	4	6	4	Base – 20%
12	12	5	5	2	Base – 20%
13	18	4	7	2	Base
14	18	3	4	3	Base – 20%
15	12	3	5	2	Base
16	18	5	6	3	Base – 20%

All choice sets included a status quo alternative, i.e., a zero price option with no environmental changes, plus two or more alternatives depending on the Design of Designs plan. Respondents were requested to choose the alternatives on the choice sets considering land use changes within a radius of about 15 km around their place of residence. At the end of the questionnaire, standard socio-demographic information was requested from respondents. For more details about the study design and sampling see [Meyerhoff et al. \(2014\)](#).

#### 4. Hypotheses and modelling approach

As we employ both stated and inferred non-attendance, the same hypotheses are investigated for both approaches. Our general assumption is that not all dimensions equally affect non-attendance. We expect that the more attributes presented on a choice set the more likely it is to observe ANA (Hypothesis 1). The reason is that respondents find it too difficult to attend to all attributes or taste heterogeneity occurs where some of the attributes are not found to be relevant for choices. Similarly, the number of alternatives is expected to influence ANA as well. Increasing the number of alternatives makes CE larger and probably more complex, thus respondents are more likely to adopt a simplifying answer heuristic (Hypothesis 2). Furthermore, we expect that ANA is positively associated with an increasing number of choice sets (Hypothesis 3). As respondents answer more choice sets they are likely to focus more on those attributes they find important and ignore the others. Regarding the number of attribute levels, we expect that the more different levels are present on a choice set the lower is ANA (Hypothesis 4). The reason is that respondents are more likely to find levels that are relevant to them. Finally, regarding attribute range we have no specific expectation regarding the relationship between either. At first glance, varying the attribute range does not change the appearance of the choice set as an increasing number of attributes or alternatives changes a choice set. [Hensher et al. \(2012\)](#) found that attribute levels and range could trigger ANA, but argue that the reason for their finding might be rather respondent-specific. Furthermore, [Hensher \(2006\)](#) found that ANA depends on the difference between experienced levels of a reference alternative and levels on other choice sets.

The relationships between both stated and inferred ANA and the design dimensions are analysed using different approaches. Firstly, we employ a non-parametric chi-squared test to determine whether there is a relationship between the design dimensions and stated non-attendance (two categorical variables). Secondly, to infer non-attendance we employ an ECLCM with pre-defined classes for attribute non-attendance. In this case we use Spearman's rank correlation coefficient to analyse the relationship between the estimated class membership probabilities for the pre-defined classes and the design dimensions as the class membership probabilities represent a continuous variable. We decided to confine the analyses of ANA to the attributes Share of Forest, Land Conversion and Cost. The reason for this is that these attributes are part of all 16 designs, thus they allow us to draw conclusions about all possible relationships. Due to the split-up according to the design plan, each of the biodiversity attributes appears only in a subset of designs.

##### 4.1. Investigating stated attribute non-attendance

After the sequence of choice tasks, we gathered information concerning the degree of stated attendance from answers to a categorical five-point scale. This scale comprised the following answer options: 'never', 'rarely', 'sometimes', 'often' and 'always'. For each attribute, respondents were requested to state their degree of attendance<sup>1</sup> (see [Scarpa et al., 2013](#), for

<sup>1</sup> Translation of the question: 'If you think back to your choice decisions, to what extent did you attend to the individual attributes of the various alternatives in your choices overall?'

a similar approach). Using this information we calculate overall stated attendance to all attributes across designs, including a graphical representation for each design dimension. Furthermore, we analyse a possible relationship between the design dimensions and stated ANA to the attributes selected for the analysis by the use of a non-parametric chi-squared test. For this we recode answers given by the respondents into binary categories. The new first category ‘not attended to’ represents the original answer options ‘never’, ‘rarely’ and ‘sometimes’ and the new second category ‘attended to’ represents the original answer options ‘often’ and ‘always’. For example, the chi-squared test then investigates the significance of the relationship between the four levels of the design dimension ‘number of choice sets’ (6, 12, 18, 24 choice sets) and two answer categories (non-attendance and attendance). Additionally, we generate scatter plots with frequencies of responses for each relationship. From these we analyse whether a change in any of the design dimensions decreases or increases the frequency of attending to an attribute.

#### 4.2. Investigating inferred attribute non-attendance

To investigate the relationship between inferred non-attendance and the design dimensions we use ECLCM with pre-defined classes. The basic idea of these models is to estimate the probability of belonging to one of the pre-defined classes<sup>2</sup> for each individual. In each of these classes, selected attributes carry zero utility weights for reflecting non-attendance. Additionally, the attribute coefficients to be estimated have been constrained to be equal across classes. Therefore, these models do not account for preference heterogeneity across respondents, i.e., the coefficient estimates are the same for each attribute across classes.

The following five classes are used in each model, i.e., for each of the 16 designs:

- For one class we assume that all attributes have been attended to: this is the full attendance class.
- We define a separate class for non-attendance to each of the attributes Share of Forest, Land Conversion, and Cost, by setting the coefficient of that attribute to zero, implying ANA.
- For the fifth class we assume that none of these three attributes has been attended to, thus all coefficients are set to zero; this is the random choice class.

The results from the ECLCM<sup>3</sup> give us the probabilities that all attributes have been attended to, one of the three investigated attributes has not been attended to, or none of the attributes has been attended to. However, in the model we did not account for response strategies ignoring any two attributes jointly. We then analyse a possible relationship between design dimensions and class probabilities by means of Spearman’s rank correlation coefficient. Additionally, we again generate scatter plots associating the probabilities of belonging to a certain class and the design dimensions. This way we graphically analyse whether a positive or negative relationship between a design dimension and the share of respondents who did not attend to an attribute in a specific design is present.

## 5. Results

### 5.1. Socio-demographic description

The Germany-wide online survey was run in December 2012. Respondents were recruited from a panel of a survey company. Each respondent was randomly allocated to one of the 16 designs. The number of respondents ranged from 79 to 222 as we previously agreed with the survey company to interview different numbers of respondents per design to account for the different numbers of choice observations resulting from each design. In total, 1684 fully completed interviews were conducted. The average interview length was 23 min and the response rate<sup>4</sup> was 29.5%. Table 3 reports the number of interviews realised and basic socio-demographics for all 16 designs. The relative shares of education refer to the highest educational level achieved.

<sup>2</sup> During the model selection process we tested various approaches. We particularly tried to account for taste heterogeneity as recent studies indicate that not accounting for taste heterogeneity may overestimate attribute non-attendance (Hess et al., 2013; Hensher et al., 2013). However, neither approach was successful. Even if it worked for one of the designs it was not possible to apply the approach successfully across all 16 designs, i.e., 16 independent samples. Firstly, an increasing number of so called ‘full-attendance classes’ were added to the equality-constrained non-attendance classes as was done, for example, by Hensher et al. (2013). The information criteria (BIC and CAIC) indicated in the majority of designs that adding more classes would be beneficial while at the same time model output could be interpreted as being less and less meaningful. For example, in many classes the cost coefficient had a positive sign and membership probabilities became very small. Secondly, we tried to add another layer of heterogeneity in the latent class model by specifying some attributes to follow a random distribution in the full attendance class. Although we tried many different model setups, these models never converged except for single data sets out of our 16 samples. Thus, neither approach was useable to infer non-attendance using the same approach across all 16 designs.

<sup>3</sup> Due to limited space, utility parameters generated in the models can be obtained from the authors upon request.

<sup>4</sup> The response rate was calculated by dividing the number of interviews completed by the number of persons invited for an interview.

**Table 3**  
Socio-demographics across designs.

Design	Interviews completed	Mean age in years	Males in %	Less than middle-school degree in %	Middle-school degree in %	High-school degree in %	University degree in %
1	85	41.98 (14.41)	44.71 (0.50)	4.71 (0.21)	29.41 (0.46)	22.35 (0.42)	43.53 (0.50)
2	82	44.32 (14.31)	41.46 (0.50)	6.10 (0.24)	24.39 (0.43)	20.73 (0.41)	48.78 (0.50)
3	124	43.81 (16.60)	42.74 (0.50)	7.26 (0.26)	25.00 (0.43)	28.23 (0.45)	39.52 (0.49)
4	80	42.50 (14.25)	50.00 (0.50)	6.25 (0.24)	32.50 (0.47)	30.00 (0.46)	31.25 (0.47)
5	104	43.30 (14.59)	54.81 (0.50)	4.81 (0.21)	24.04 (0.43)	25.00 (0.44)	46.15 (0.50)
6	82	42.32 (13.21)	46.34 (0.50)	7.32 (0.26)	29.27 (0.46)	18.29 (0.39)	45.12 (0.50)
7	222	42.49 (13.12)	46.85 (0.50)	5.41 (0.23)	31.08 (0.46)	27.48 (0.45)	36.04 (0.48)
8	79	42.54 (12.75)	56.96 (0.50)	10.13 (0.30)	25.32 (0.44)	29.11 (0.46)	35.44 (0.48)
9	81	43.47 (14.06)	37.04 (0.50)	4.94 (0.22)	28.40 (0.45)	30.86 (0.48)	35.80 (0.48)
10	151	42.02 (13.69)	49.67 (0.50)	10.60 (0.31)	20.53 (0.41)	29.80 (0.46)	39.07 (0.49)
11	88	41.44 (13.81)	56.62 (0.50)	4.55 (0.21)	36.36 (0.48)	21.59 (0.41)	37.50 (0.49)
12	81	39.47 (13.29)	39.51 (0.50)	3.70 (0.19)	22.22 (0.42)	38.27 (0.49)	35.80 (0.48)
13	113	42.09 (13.15)	55.75 (0.50)	4.42 (0.21)	25.66 (0.44)	26.55 (0.44)	43.36 (0.50)
14	84	43.40 (12.42)	0.50 (0.50)	15.48 (0.36)	23.81 (0.44)	23.81 (0.43)	36.90 (0.49)
15	144	40.44 (13.60)	0.50 (0.50)	7.64 (0.27)	29.86 (0.46)	26.39 (0.44)	36.11 (0.48)
16	84	42.15 (13.80)	48.81 (0.50)	5.95 (0.24)	23.81 (0.43)	28.57 (0.45)	41.67 (0.50)
Total	1684	42.33 (13.59)	48.34 (0.50)	6.83 (0.25)	27.08 (0.44)	26.84 (0.44)	39.25 (0.49)

Note: Standard deviations in parentheses.

**Table 4**  
Stated attendance to all attributes (percentage of respondents who answered 'often' or 'always' attended to).

Design	9	6	8	12	14	5	13	16	15	1	4	11	3	7	10	2
Stated Att.	32	24	20	20	19	17	17	17	15	13	13	13	10	9	9	9

## 5.2. Overall stated attribute non-attendance across designs

As a first step in our analysis, we calculated the share of respondents who stated that they 'often' or 'always' attended to all attributes presented in the choice sets. Table 4 presents these results in descending order of stated attribute attendance per design.

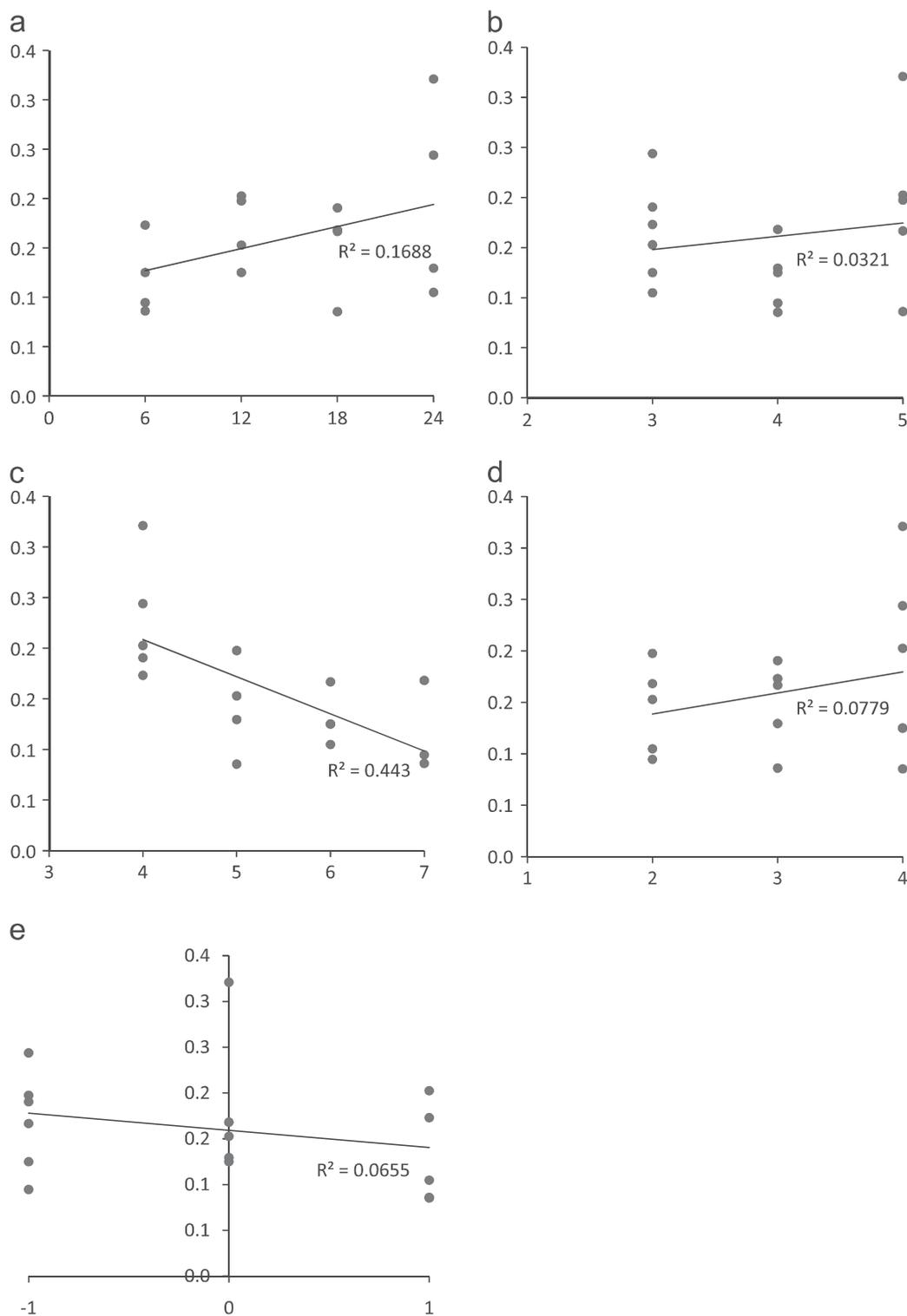
Fig. 1 presents the results from Table 4 separately for the five design dimensions: number of choice sets, number of alternatives, number of attributes, number of levels and level range in ascending order of the levels of each design dimension. A linear trend line is added each time. As can be seen, only in graph c of Fig. 1a clear negative relationship between stated non-attendance and the design dimension is visible: attendance is higher the fewer attributes a design comprises. This is in line with our first hypothesis that a lower number of attributes is associated with higher stated attribute attendance. The other four plots do not show clear positive or negative relationships. The preliminary graphical analysis leads, therefore, to the conclusion that the number of choice sets, the number of alternatives, the number of levels and the level range does not have an influence on the stated attendance to attributes.

## 5.3. Attribute specific stated non-attendance across designs

The responses to the non-attendance scale are reported in Table 5. The mean values, calculated from response scores for each attribute separately, indicate the degree of ANA to each single attribute in each design. Values closer to five indicate higher levels of attendance while values closer to one indicate higher levels of non-attendance.

The attribute Share of Forest receives rather high stated attendance values (almost always > 4 and mean 4.10, meaning it was 'often attended to') whereas the stated attendance values for other attributes are relatively lower (almost always < 4). Attendance to Land Conversion in some designs also lies around 4, but mean attribute attendance over all designs is 3.75. Attendance to the Cost attribute is 3.29 meaning it was 'sometimes attended to'; this effect is visible in all designs. This finding is in line with other literature. Campbell et al. (2008) found in their study that Cost was taken into account by 69% of respondents, less than all other attributes presented.

Next, we use a chi-squared test to investigate the relationship between two categorical variables, the five design dimensions and the two recoded stated answer categories to single attributes across designs (non-attendance and attendance). From scatter plots we aim to deduce the direction of the influence of the design dimensions on stated attendance. Fig. 2 presents *p*-values from the chi-squared tests (significant correlations are bold) and corresponding scatter plots. As the results reveal, for the majority of attributes we cannot find significant or clear relationships between stated attendance and the design dimensions. Nevertheless, for some combinations of attributes and design dimensions this relationship is significant. A significant negative effect on stated attendance to Land Conversion was detected for the number



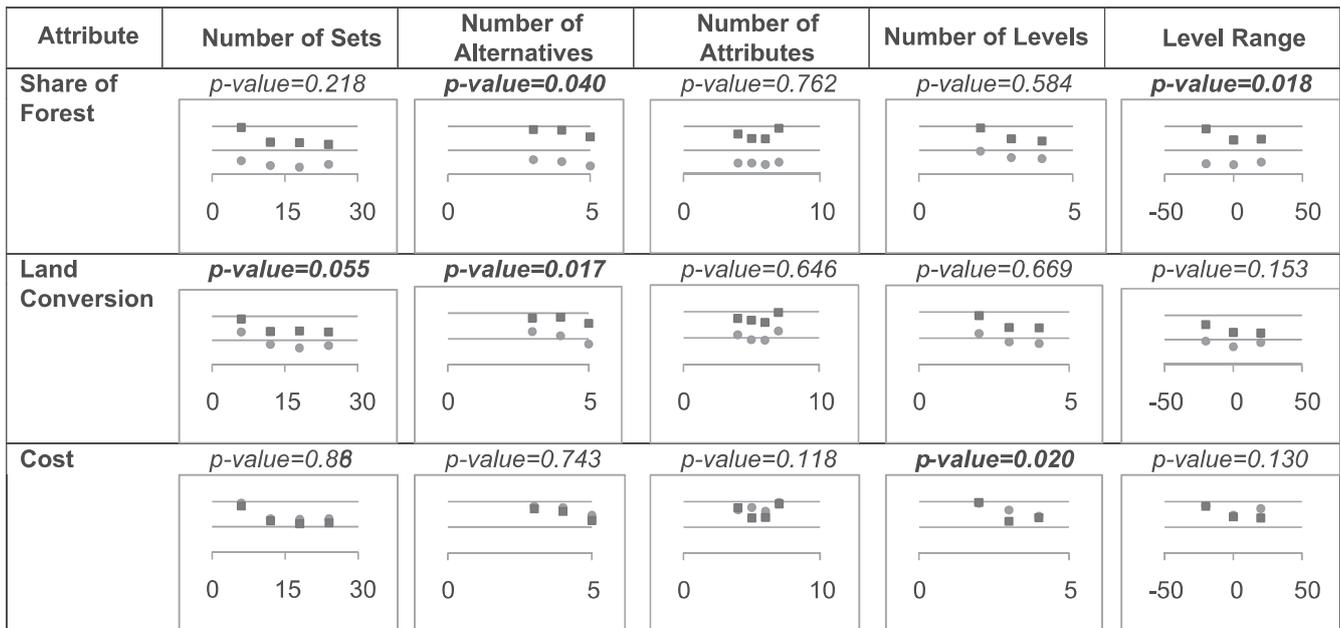
**Fig. 1.** Stated attribute attendance per design dimension. (a) Number of choice sets. (b) Number of alternatives. (c) Number of attributes. (d) Number of levels. (e) Level range. Legend: x-axis—levels of design dimension. y-axis—stated attendance (in decimal numbers). Grey dots—Stated attendance. Black lines—Linear (Stated attendance).

of choice sets presented. This means that the more choice sets are presented the less likely it becomes that a respondent states that he/she has attended to the attribute. The number of alternatives has a significant and overall negative influence on stated attendance to Share of Forest and Land Conversion. For the number of attributes no significant relation was detected. For the number of levels included in the designs, we observe a negative and significant effect on stated attendance to Cost. Interestingly, this effect occurs despite the fact that for the Cost attribute neither the number nor the range of levels (i.e., the price) varies. A significant negative effect was also observed for the influence of the level range on Share of Forest. If the level range becomes wider, less respondents answer that they attended to Share of Forest.

**Table 5**  
Means of attribute attendance values coded on a five-point scale.

Attributes	Design																Mean
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Share of Forest	4.16	3.99	3.87	4.03	3.90	3.91	4.13	4.01	4.26	4.15	4.20	4.35	4.20	4.22	3.93	4.36	4.10
Land Conversion	3.82	3.75	3.65	3.64	3.50	3.52	3.65	3.58	4.01	3.73	3.76	3.97	3.89	3.85	3.64	4.05	3.75
Cost	3.06	3.01	3.11	3.23	3.53	3.50	3.38	3.22	3.45	3.03	3.38	3.47	3.52	3.23	3.25	3.20	3.29

Note: Number of observations for Share of Forest, Land Conversion and Cost: N=1684.



**Fig. 2.** Chi-squared tests and scatter plots. Note: Correlations significant at the 10% level or higher are bold; the units of the y-axis are not presented to have a larger scatter plot area. x-axis: levels of design dimension; y-axis: frequencies of responses on five-point scale in binary coding; black square: 'attended to', grey dot: 'not attended to'.

5.4. Inferred attribute non-attendance across designs

Table 6 reports class membership probabilities of ANA inferred from the ECLCM. The last column reports the mean, minimum and maximum probabilities across all designs. The attribute with the highest average probability of ANA is the Cost attribute (average 27.2%). The maximum predicted class size for non-attendance to Cost is 40.1% (Design 10) while the minimum predicted class size is 10.8% (Design 3). Adding the probability of not attending to any attribute (class 'Random choice') implies that approximately 41.9% of respondents may not have attended to the Cost attribute. Next, the attribute Share of Forest has a mean class probability of ANA of 18.1% (class 'Share of Forest not attended to'). It varies from 37.4% (Design 3) to 8.5% (Design 16). Again, adding the probability of not attending to any attribute results in a probability of ANA to Share of Forest of 32.8%. The third non-attendance class investigated here, non-attendance to Land Conversion, has a mean class probability of 17.0% across designs with the highest probability equal to 29.4% (Design 11) and the lowest probability equal to 8.2% (Design 10). The overall probability after adding the mean membership probability for the class 'Random choice' results in a probability of ANA of 31.7%. Finally, across all designs respondents have an average probability of attending to all attributes of 23.1%, ranging from 17.1% (Design 12) to 33.0% (Design 14).

Next, we turn to the relationship between the design dimensions of the 16 samples and the probabilities to attend to or not attend to attributes according to the pre-defined classes.

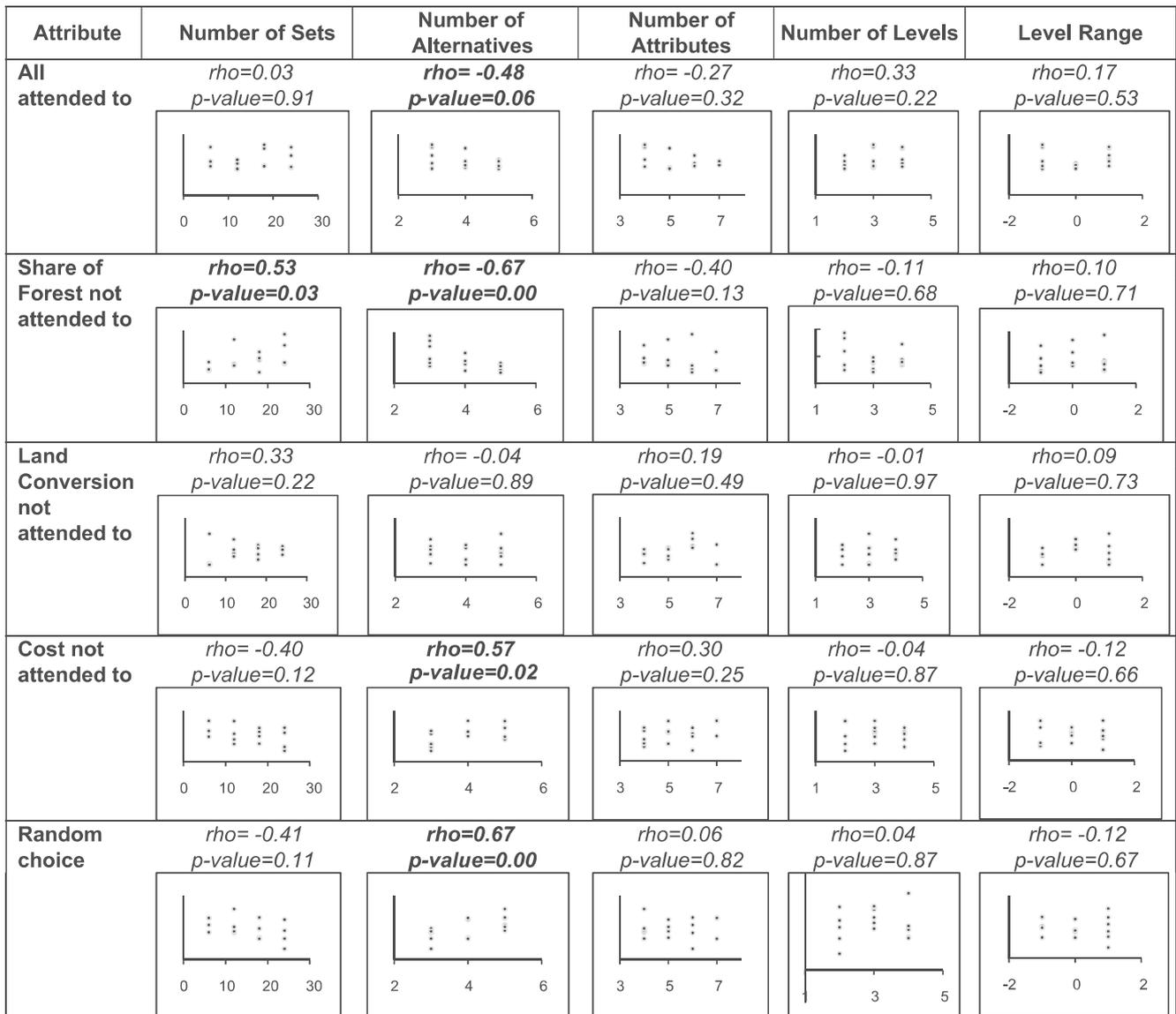
Spearman's rank correlation coefficient was used to analyse the relationship between single design dimensions and class membership probabilities. Fig. 3 presents the correlation coefficients together with the associated p-values and scatter plots for each combination of analysed design dimensions and classes included in the ECLCM. These scatter plots should help to infer the direction of a possible relationship, i.e., whether an increasing design dimension affects inferred ANA positively or negatively.

It becomes obvious from Fig. 3 that only one of the design dimensions has a systematic influence on inferred attribute non-attendance. The number of alternatives in four out of five cases significantly influences the probabilities of attending to or not-attending to an attribute. The probability to attend to all attributes decreases with an increasing number of alternatives, which is in line with our expectations that with an increasing dimensionality the probability of ANA also

**Table 6**  
Class membership probabilities in %.

Attendance class	Design															Mean Min Max			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14		15	16	
All attended to	17.9	30.7	25.9	20.8	31.5	31.5	22.1	23.2	18.5	22.1	18.9	17.1	19.5	33.0	33.0	17.2	19.0	23.1	
	?(4.2)	?(5.3)	?(4.0)	?(5.5)	?(5.9)	?(5.3)	?(4.1)	?(5.1)	?(4.3)	?(4.1)	?(4.9)	?(4.3)	?(4.0)	?(5.3)	?(5.3)	?(3.5)	?(4.3)	17.1	
Share of Forest not attended to	15.1	17.7	37.4	13.6	16.2	29.0	10.0	14.9	15.9	10.0	10.7	13.7	23.9	19.3	33.5	33.5	8.5	18.1	
	?(3.9)	?(4.4)	?(4.4)	?(4.5)	?(4.5)	?(5.1)	?(4.0)	?(4.4)	?(4.1)	?(4.0)	?(4.1)	?(3.8)	?(4.0)	?(4.3)	?(4.1)	?(3.0)	?(3.0)	8.5	
Land Conversion not attended to	19.7	11.8	20.8	25.6	9.1	15.0	8.2	16.2	18.5	8.2	29.4	13.9	21.7	15.3	18.5	18.5	19.5	17.0	
	?(4.5)	?(3.8)	?(3.8)	?(5.4)	?(3.8)	?(4.1)	?(4.1)	?(4.7)	?(4.3)	?(4.1)	?(6.0)	?(4.3)	?(4.1)	?(4.0)	?(3.6)	?(3.6)	?(4.7)	8.2	
Cost not attended to	28.5	29.5	10.8	27.4	30.3	14.5	40.1	21.9	33.4	40.1	24.7	39.9	25.1	17.7	17.4	17.4	33.2	27.2	
	?(5.1)	?(5.3)	?(3.0)	?(5.6)	?(5.7)	?(4.2)	?(6.2)	?(4.7)	?(5.2)	?(6.2)	?(5.8)	?(5.7)	?(4.4)	?(4.5)	?(3.6)	?(3.6)	?(5.3)	10.8	
Random choice	18.8	10.3	5.1	12.6	12.9	9.9	19.6	23.9	13.7	19.6	16.4	15.4	9.8	14.7	13.3	13.3	19.8	14.7	
	?(4.5)	?(3.5)	?(2.2)	?(4.5)	?(4.6)	?(3.7)	?(2.9)	?(5.1)	?(3.8)	?(2.9)	?(5.5)	?(4.4)	?(2.9)	?(4.0)	?(3.1)	?(3.1)	?(4.8)	5.1	
																			23.9

Note: Standard errors in parenthesis.



**Fig. 3.** Spearman's rank correlation coefficient. Note: Correlations significant at the 5% level or higher are bold; the units of the y-axis are not presented to have a larger scatter plot area. x-axis: levels of design dimension; y-axis: class membership probabilities from Table 6.

increases. Interestingly, the probability to attend to the attribute Share of Forest increases with an increasing number of alternatives. For the Cost attribute we see that with an increasing number of alternatives the probability of not attending to Cost increases. Also, and also in line with our expectations, the probability of random choice increases when choice sets comprise more alternatives. Furthermore, the only relationship that is statistically significant is between the number of sets and the non-attendance to the attribute Share of Forest. According to the scatter plot this relationship is positive, which is in line with our expectation that an increasing number of choice sets positively influences the probability of not attending to this attribute. For the other relationships we could not find significant associations. Thus, overall the results indicate that with an increasing number of alternatives respondents seem to focus on the two attributes Share of Forest and Land Conversion while paying less attention to the other attributes, especially the Cost attribute.

## 6. Discussion

The aim of this study was to investigate the relation between the dimensionality of a CE and ANA. To determine the influence of these dimensions we applied a Design of Designs approach resulting in 16 different treatments. Stated non-attendance was recorded using a five-point non-attendance response scale while for inferring non-attendance we used an ECLCM.

The results from the analysis of stated ANA show that the attribute Share of Forest was favoured by many respondents, i.e., they attended to it to a large extent across all designs ('often attended to'). Stated attendance to the Cost attribute was rather low ('sometimes attended to'). Mean attendance to Land Conversion lay between the two.

We investigated the relation between stated attendance to single attributes and design dimensions from chi-squared tests, which revealed only five statistically significant relations. Scatter plots derived from frequencies of responses showed that higher numbers of choice sets tend to negatively influence the frequency of stating ‘attendance to’ an attribute (e.g., to the attribute Land Conversion). A higher number of alternatives tends to negatively influence attendance to attributes (e.g., to the attribute Share of Forest or Land Conversion). Also, a higher number of levels negatively influenced the probability of stating ‘attendance to’ the Cost attribute, also a wider range of levels negatively influenced stated attendance (e.g., for Share of Forest).

Inferred ANA was derived from an ECLCM, which estimates class membership probabilities for pre-defined classes (‘All attended to’, ‘Share of Forest not attended to’, ‘Land Conversion not attended to’, ‘Cost not attended to’, ‘Random choice/No attribute attended to’). These showed that attendance differs quite substantially between the classes. Particularly worth mentioning is that – according to our latent class model – only a relatively small proportion of respondents considered all attributes. The class size for this group of respondents varied between 17 and 33%. On the other hand, the probabilities for random choice varied between 5 and 24% indicating that respondents allocated to some designs have a significant probability to not consider any attribute.

Using Spearman’s rank correlation coefficient we deduced that higher numbers of choice sets positively influenced the probability of non-attending to Share of Forest. Higher numbers of alternatives significantly influenced four out of five class membership probabilities investigated: probabilities of ‘all attributes attended to’ and ‘Share of Forest not attended to’ classes decreased with more alternatives, but ‘Cost not attended to’ and ‘Random choice’ classes increased with more alternatives.

Regarding attendance to Cost, we found that a large share of the respondents scarcely attended to Cost even though attention to the Cost attribute is crucial from an economic point of view to calculate marginal willingness to pay and welfare measures. This result is visible in both stated and inferred attendance: Respondents stated that it was ‘sometimes’ included in the decision and class membership probabilities for ‘Cost not attended to’ lay between 11 and 40%, meaning that a substantial number of respondents did not include the Cost attribute in their decision. Similar results were found in other studies (e.g., see [Campbell et al., 2008](#), for inferred results; [Hensher et al., 2012](#); [Kragt, 2013](#)). Reasons for ANA to Cost could be that as long as costs lie within a certain range, they are not considered important. We know this from focus group discussions where participants indicated that they don’t care about costs unless they are beyond a certain threshold. For designs with a high number of attributes, ANA to Cost may additionally be driven by more important attributes being offered, which attract more attention than the less important Cost attribute. We believe that social desirability may also have affected stated non-attendance to Cost. It may be that respondents wanted to conceal their true cost awareness when asked about attribute attendance, thus they stated less attendance to Cost than actually present.

Looking at our results, the first, and probably most important, conclusion is that in both approaches – stated and inferred – the relationship between attribute attendance and design dimensions is not very strong. Nevertheless, when going into detail, our findings regarding stated and inferred ANA suggest some relationships between the design dimensions and ANA. The results from stated ANA indicate that a higher number of alternatives are significantly influencing this form of attendance, and the results from inferred ANA also indicate that the number of alternatives is the main driver behind non-attendance. Regarding our hypotheses the analysis found no significant influence of the number of attributes on ANA, which does not confirm our first hypothesis that the number of attributes is a driver of non-attendance. However, the second hypothesis, stating that an increasing number of alternatives influences ANA, has been supported by the results from the stated and the inferred approach. Also, the expected direction of the effect for an increase in the number of alternatives has been found for the majority of classes (fewer ‘all attended to’, more ‘Cost not attended to’, more ‘Random choice’). An increasing number of alternatives could make comparisons, and subsequently the choosing of an alternative, more difficult. In the end this might lead to ignoring attributes as a response to the higher level of difficulty. The results of the two approaches are consistent with the third hypothesis, stating that an increasing number of choice sets may increase non-attendance. From the stated ANA analysis we deduced that Land Conversion was attended to less when more choice sets were presented. From the inferred analysis we deduced that more choice sets led to more ANA to Share of Forest. It may be that respondents perceived a higher number of sets as exhausting and just focused on some attributes. The fourth hypothesis, stating that more levels lead to more attendance, was not supported by the stated analysis. Here, more levels led to less attendance to the Cost attribute.

Asking respondents to state their level of attendance has been a subject of debate, i.e., some researchers question whether respondents correctly state to what extent they have or have not attended to certain attributes. However, from the literature it is not clear whether models inferring ANA really do better. The ECLCMs in particular are suspected to overestimate ANA because they do not account for taste heterogeneity ([Hess et al., 2013](#); [Hensher et al., 2013](#)). Therefore, our models might overestimate the degree of ANA and should be interpreted with a degree of caution. On the other hand, as outlined in [Section 4](#), we estimated various models incorporating taste heterogeneity via additional full attendance classes or additional random parameters within classes, but they were not informative or sufficiently interpretable. This is especially true as we tried to apply a consistent approach across all 16 designs, i.e., samples. It may have been possible to adjust individual models to each sample, this would, however, have required a very high effort. [Hensher et al. \(2013\)](#) report in a footnote that a huge amount of time was necessary to identify what they thought of as an appropriate model. To do this for all 16 samples would have been beyond our resources and based on our experience with these models, we think that it is questionable whether the insights gained would have justified this effort.

To conclude, from what we found in this study, results indicate that the connection between dimensionality and non-attendance to attributes is not very strong. This is true for both stated and inferred ANA. Nevertheless, based on our results, we suggest that a negative effect on attribute attendance should be expected from the number of alternatives and the number of choice sets. In contrast, we would not expect that a higher number of attributes, levels and level range results in higher degrees of ANA in future CE. Future research options in this field, apart from other model specifications or the application of the RPANA model (Collins et al., 2013), could be to consider other attribute processing rules such as the aggregation of common-metric attributes. For this analysis, the biodiversity attributes could be reconsidered as this decision rule might be used for these attributes.

## Acknowledgements

We would like to thank two anonymous reviewers for their very helpful comments. Also, we would like to thank the editor, David Hensher, for his support. The authors acknowledge financial support from the Basque Government through grant IT-642-13 and from UPV/EHU through grant US12/09. Furthermore, funding for this research was provided by the German Federal Ministry of Education and Research (Fkz. 033L029G; Fkz 01LL0909A) and is gratefully acknowledged.

## Appendix

See [Table A1](#)

**Table A1**

Literature regarding attribute non-attendance.

Study	Stated ANA		Inferred ANA		Key findings
	Model	Treatment in RUM	Model	Treatment in RUM	
Hensher et al. (2005)	Mixed Logit	Individual-level 0-utility weights			Higher WTP when not accounting for ANA
Hensher (2006)	Mixed Logit	Individual-level 0-utility weights			Lower WTP when accounting for ANA
Hensher et al. (2007)	Mixed Logit	Individual-level 0-utility weights, expected maximum utility			Higher WTP when accounting for ANA
Campbell et al. (2008)	MECLM	Individual-level 0-utility weights, parameterisation of the scale parameter based on attribute processing strategies			Better model fit, lower WTP and different scale parameters when accounting for ANA
Hensher and Rose (2009)	Mixed Logit	Individual-level 0-utility weights			Better model fit and higher WTP when accounting for ANA
Scarpa et al. (2009)			ECLCM; STAS	0-Utility weights (a priori); 0-weight (from data)	Better model fit and lower WTP when accounting for ANA
Carlsson et al. (2010)	Mixed Logit	Individual-level 0-utility weights			No significant differences in preferences and WTP in samples accounting and not accounting for stated ANA
Hensher and Greene (2010)			ECLCM	0-Utility weights	Higher WTP when accounting for ANA
Hess and Hensher (2010)	MNL; Mixed Logit	Separate parameters for attendance and non-attendance groups		Coefficient of variation for the conditional distribution	Results from inferred and stated ANA are not consistent, better fit for the inferred approach
Scarpa et al. (2010)	MNL	Individual-level and choice set specific (choice-set ANA) 0-utility weights			Better model fit and more efficient WTP estimates when accounting for choice-set ANA
Balcombe et al. (2011)	Mixed Logit	Unrestricted and restricted interactions, individual-level 0-utility weights			Better model fit when accounting for ANA
Campbell et al. (2011)			Scale-adjusted ECLCM	0-Utility weights	Lower WTP when accounting for ANA
Hensher et al. (2012)			ECLCM		ANA and WTP also depend on relevance of levels and range to respondents
Rose et al. (2012)	Mixed Logit	Ignored attributes excluded from estimation procedure			Significant differences in parameter estimates when accounting for ANA

Table A1 (continued)

Study	Stated ANA		Inferred ANA		Key findings
	Model	Treatment in RUM	Model	Treatment in RUM	
Collins et al. (2013)			(Generalised) RPANA		Generalised RPANA links stated and inferred ANA, better performance than existing models
Colombo et al. (2013)	Mixed Logit	Separate parameters for attendance and non-attendance groups			Accounting for 'sometimes attendance' offers advantages over only 2 attendance groups
Hensher et al. (2013)			Random parameter LCM	0-Parameters (mean and SD)+class heterogeneity	Accounting for ANA gives more model fit than random parameters, random parameters increase probability of full-attendance
Hess et al. (2013)			ECLCM; LCM; LC-MMNL	Coefficients set to zero/estimated (point value, continuous distribution)	Better model fit in combined LC-MMNL, implied rate of ANA reduced
Hole et al. (2013)	EAAModel; MEAA	Indicator variable		Attribute not-attended to constrained to 0	Better fit but similar WTP when accounting for ANA
Kragt (2013)	MXL; Mixed Logit	Individual-level 0-utility weights	ECLCM	0-Utility weights	Best fit for the LCM, little concordance between stated and inferred ANA
Lagarde (2013)			ECLCM	0-Utility weights	Better model fit and higher WTP when accounting for ANA, no difference for predicted probabilities
Scarpa et al. (2013)	MXL; MNL	Separate parameters for attendance and non-attendance groups	ECLCM	0-Utility weights	Best fit for the ECLCM

Note: MECLM=multinomial error component logit model; (M)EAAModel=(mixed) endogenous attribute attendance model (Hole, 2011); STAS=stochastic attribute selection; MXL=random parameter model with error components; LC-MMNL=latent class mixed logit model.

## References

- Balcombe, K., Burton, M., Rigby, D., 2011. Skew and attribute non-attendance within the Bayesian mixed logit model. *J. Environ. Econ. Manage.* 62, 446–461.
- Boxall, P., Adamowicz, W.L., Moon, A., 2009. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Aust. J. Agric. Res. Econ.* 53, 503–519.
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.
- Campbell, D., Hensher, D.A., Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. *J. Environ. Plann. Manage.* 54 (8), 1061–1076.
- Campbell, D., Hutchinson, W.G., Scarpa, R., 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environ. Res. Econ.* 41 (3), 401–417.
- Carlsson, F., Kataria, M., Lampi, E., 2010. Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environ. Res. Econ.* 47, 65–89.
- Caussade, S., de Dios Ortuzar, J., Rizzi, L.L., Hensher, D.A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transp. Res. Part B* 39, 621–640.
- Collins, A.T., Rose, J.M., Hensher, D.A., 2013. Specification issues in a generalised random parameters attribute nonattendance model. *Transp. Res. Part B* 56, 234–253.
- Colombo, S., Christie, M., Hanley, N., 2013. What are the consequences of ignoring attributes in choice experiments? Implications for ecosystem service valuation. *Ecol. Econ.* 96, 25–35.
- Hensher, D.A., 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. *J. Transp. Econ. Policy* 38 (3), 425–446.
- Hensher, D.A., 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. *J. Appl. Econometrics* 21, 861–878.
- Hensher, D.A., Collins, A.T., Greene, W.H., 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. *Transportation* 40 (5), 1003–1020.
- Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical Econ.* 39 (2), 413–426.
- Hensher, D.A., Rose, J.M., 2009. Simplifying choice through attribute preservation or non-attendance: implications for willingness to pay. *Transp. Res. Part E* 45, 583–590.
- Hensher, D.A., Rose, J.M., Bertoia, T., 2007. The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies. *Transp. Res. Part E* 43, 73–89.
- Hensher, D.A., Rose, J.M., Greene, W.H., 2005. The implications on willingness to pay of respondents ignoring specific attributes. *Transportation* 32, 203–222.
- Hensher, D.A., Rose, J.M., Greene, W.H., 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation* 39, 235–245.
- Hess, S., Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transp. Res. Part B* 44 (6), 781–790.

- Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., Caussade, S., 2013. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation* 40 (3), 583–607.
- Hole, A.R., 2011. A discrete choice model with endogenous attribute attendance. *Econ. Lett.* 110, 203–205.
- Hole, A.R., Kolstad, J.R., Gyrd-Hansen, D., 2013. Inferred vs. stated attribute non-attendance in choice experiments: a study of doctors' prescription behaviour. *J. Econ. Behav. Organiz.* 96, 21–31.
- Kragt, M.E., 2013. Stated and inferred attribute attendance models: a comparison with environmental choice experiments. *J. Agric. Econ.* 64 (3), 719–736.
- Lagarde, M., 2013. Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health Econ.* 22, 554–567.
- Meyerhoff, J., Oehlmann, M., Weller, P., 2014. The influence of design dimensions on stated choices in an environmental context. *Environ. Res. Econ.* <http://dx.doi.org/10.1007/s10640-014-9797-5>.
- Rose, J.M., Hensher, D.A., Greene, W., Washington, S., 2012. Attribute exclusion strategies in airline choice: accounting for exogenous information on decision maker processing strategies in models of discrete choice. *Transportmetrica* 8 (5), 344–360.
- Scarpa, R., Gilbride, T.J., Campbell, D., Hensher, D.A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *Eur. Rev. Agric. Econ.* 36 (2), 151–174.
- Scarpa, R., Rose, J.M., 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Aust. J. Agric. Res. Econ.* 52, 253–282.
- Scarpa, R., Thiene, M., Hensher, D.A., 2010. Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter? *Land Econ.* 86 (4), 817–839.
- Scarpa, R., Zanolli, R., Bruschi, V., Naspetti, S., 2013. Inferred and stated attribute non-attendance in food choice experiments. *Am. J. Agric. Econ.* 95 (1), 165–180.

4. Meyerhoff, J., Oehlmann, M., und Weller, P., 2015. The influence of design dimensions on stated choices in an environmental context. *Environmental and resource economics*, 61 (3), 385-407.

# The Influence of Design Dimensions on Stated Choices in an Environmental Context

- Final accepted manuscript -

Jürgen Meyerhoff \*, Technische Universität Berlin, Institute for Landscape and Environmental  
Planning, Straße des 17. Juni 145, D-10623 Berlin, Phone # +49 30 31473322, [juergen.meyerhoff@tu-berlin.de](mailto:juergen.meyerhoff@tu-berlin.de)

\* corresponding author

Malte Oehlmann, Technische Universität Berlin, Institute for Landscape and Environmental Planning  
Straße des 17. Juni 145, D-10623 Berlin, [malte.oehlmann@tu-berlin.de](mailto:malte.oehlmann@tu-berlin.de)

Priska Weller, Johann Heinrich von Thünen Institute, Institute of Forest Economics, Leuschnerstraße  
91, D-21031 Hamburg, [priska.weller@ti.bund.de](mailto:priska.weller@ti.bund.de)

## 1. Introduction

Discrete choice experiments (CE) are widely used to elicit consumer preferences in various applications including marketing, health economics, transportation and environmental valuation. CEs usually consist of a series of hypothetical scenarios or choice sets, each including two or more alternatives. These alternatives are described by attributes whose levels are varied systematically. For each choice set respondents are asked to choose their preferred alternative (Louviere et al. 2000). The number of choice sets, alternatives, and attributes as well as the number of levels and their range are the design dimensions of a CE. These are to be specified by the researcher. The dimensionality of a CE can vary significantly across different studies. For instance, in the majority of CE studies respondents are offered four to ten choice sets and two to four alternatives. However, there are occasions when participants are asked to assess up to 26 choice tasks (Czajkowski et al. 2012) and choose among 12 or more alternatives (Chung et al. 2011).

Assuming neoclassic economic theory, which suggests that decision makers can perfectly and consistently process information, the design dimensionality should not influence choice outcomes (Swait and Adamowicz 2001). As a consequence, the complexity of the CE, the ability of the individual to make complex decisions and the effect of the choice context on decision strategies are often not considered in statistical models. However, many authors have highlighted the need to account for the limited ability of individuals to process complex information. Heiner (1983), for instance, argued that choice complexity can influence choice consistency and the more complex the choice task, the higher the gap between individual's cognitive ability and cognitive demand. This result leads to a trade-off to be made by the researcher. On the one hand, one might be tempted to increase task complexity, for instance, to increase the number of observations or to gather more information by increasing the number of attributes. On the other hand, this comes at the cost of higher cognitive burden and the risk of significantly distorted model estimates (DeShazo and Fermo 2002). As a result, there is so far no agreement in the literature on the optimal task complexity in CEs.

Task complexity can be viewed as a part of the unobserved factors influencing choice outcomes. Generally, unobserved influences are classified as omitted variables, measurement errors in the observed attributes and alternatives, true task complexity that imposes variation in cognitive difficulty, uncertainty attributable to many sources such as stimulus ambiguity and beliefs about future states and peer impacts (Hensher 2006).

Our study investigates the issue of true task complexity by varying five design dimensions. The research is largely motivated by a series of studies investigating the influence of five design dimensions - the number of choice sets, alternatives, attributes, attribute levels and their range - on CE outcomes in the context of transportation (Hensher 2004; Caussade et al. 2005; Hensher 2006; Rose et al. 2009). In these studies, 16 different treatments were used which were generated by a Design of Designs (DoD) approach originally introduced by Hensher (2004). The attributes used considered different travel times and costs. The present case study is based on a nation-wide online survey carried out in Germany about land use changes. The survey was completed in December 2012 and incorporated aspects of land conversion, the share of forest and different biodiversity attributes.

Similar to Hensher (2004) and Caussade et al. (2005) we used 16 different split samples. To our knowledge, this is the first study applying a DoD approach in environmental valuation by systematically varying five design dimensions. We analyse the influence of the five design dimensions in terms of two aspects. First, we investigate the relationship between task complexity and participants' drop-out behaviour by using descriptive statistics and by specifying a binary logit model. Second, we estimate a joint multinomial logit model (MNL) and heteroskedastic logit models (HL) with the scale parameter specified as a function of the design dimensions and socio-demographic characteristics of the respondents. This allows us to determine the impact of the five design dimensions on the error term variance which is inversely related to the scale parameter. Willingness-to-pay (WTP) estimates are subsequently obtained from both models and compared to each other. In the remainder Section 2 outlines previous literature and the hypotheses to be tested. Next, Section 3

presents the modelling approach before detailing the study design and implementation (Section 4) and presenting results (Section 5). Finally, Section 6 discusses the results.

## **2. Literature Review and Hypotheses**

The influence of task complexity in stated CEs has been investigated in several studies. Typically, complexity issues have been analysed in the context of health economics (Ryan and Wordsworth 2000; Ratcliffe and Longworth 2002; Bech et al. 2011), marketing research (Dellaert et al. 1999; Dellaert et al. 2012) and transportation (Hensher et al. 2001; Hensher 2004; Caussade et al. 2005; Hensher 2006; Rose et al. 2009). The research has focused largely on dimensionality influences on the error variance or scale parameter considering the effects of fatigue and learning (Dellaert et al. 1999; DeShazo and Fermo 2002; Arentze et al. 2003; Caussade et al. 2005; Rolfe and Bennett 2009; Chung et al. 2011; Czajkowski et al. 2012) and on WTP estimates (Ryan and Wordsworth 2000; Hensher 2004; Hensher 2006; Bech et al. 2011; McNair et al. 2011). Effects on attribute weights (Arentze et al. 2003), response and completion rate (Hensher et al. 2001; Bech et al. 2011; Louviere et al. 2013), decision time (Dellaert et al. 2012) and perceived choice certainty (Rose et al. 2009; Brouwer et al. 2010; Bech et al. 2011) have also been analysed in existing research. Regarding the impacts on the error variance and in the light of studies published in recent years, we largely follow the hypotheses that were put forward by Caussade et al. (2005) and compare their results to our findings in the context of environmental valuation.

### *2.1. Task Complexity and Drop-out Rates*

To our knowledge only very few studies have investigated the impact of task complexity on participants' drop-out behaviour in discrete CEs. Two studies explicitly investigate the relationship between the number of choice sets and the response rate. Hensher et al. (2001) found little differences in response rates in mail surveys with varying numbers of choice tasks, while Bech et al. (2011) found no significant influence in an online survey. However, it

has to be noted that response and drop-out rates must be distinguished from each other. Whereas the former gives the ratio between those respondents who completed the questionnaire and those contacted, the latter shows the percentage of those participants who prematurely abandoned the questionnaire (Vicente and Reis 2010). So far the most comprehensive study concerning completion rates was presented by Louviere et al. (2013). They used a web panel to elicit preferences for two consumer goods, pizza and flights. Across interviews the authors systematically varied the number of choice sets (16 or 32), the number of alternatives (3 to 5), the number of attributes (6 to 12) and the type of statistical design ("Street and Burgess" or "Kuhfeld" design). They found that, among other things, completion rates are relatively unaffected by presenting more choice sets to respondents, but decline when the sets comprise more alternatives. In contrast, studies carried out in other disciplines concluded that the survey length has a significant influence on drop-out rates (see for example Vicente and Reis 2010). Furthermore, a study by Galesic (2006) shows that the lower the experienced burden, the lower the risk to drop out. We thus expect that both, the number of choice sets and the number of alternatives, are positively related to drop-out rates although the results presented by Louviere et al. (2013) indicate only a weak effect.

Finally, for the number of attributes as well as for the number of levels and their range, it may be argued that the number of comparisons to be made increases with both, more attributes and a higher number of levels. Moreover, comparisons between alternatives might be easier to assess for attribute levels which have a narrow range (Caussade et al. 2005). As a result, we also hypothesise a positive relationship between these design dimensions and drop-out rates.

## 2.2. *Task Complexity and Error Variance*

*Number of Choice Sets:* So far the most frequently investigated design dimension is probably the number of choice sets. However, there is still no consensus in the literature about impacts on the error variance. On the one hand, Hensher et al. (2001) found little evidence

for fatigue effects for even 32 choice sets; Czajkowski et al. (2012) support this inference. On the other hand, Bradley and Daly (1994) found the error variance to increase as the number of choice sets grew. Caussade et al. (2005) and Chung et al. (2011) observed a U-shaped relationship with the error variance decreasing up to a threshold (nine/ten sets in Caussade et al. 2005; six sets in Chung et al. 2011) and then increasing beyond this. A similar pattern was found by Bech et al. (2011). They argue that the error variance initially decreases due to learning effects while subsequent fatigue effects cause the error variance to increase. Also, findings by Scarpa et al. (2011) point in this direction, who based their analysis on rank-ordered data elicited with the best-worst approach. In their study each respondent faced 16 choice sets. Scale increased gradually from the first to the eleventh set and declined for the last three sets. Following these arguments we expect a U-shaped pattern for the relation between the number of choice sets and scale.

*Number of Alternatives:* The evidence and subsequent conclusions about the impact of the number of alternatives tend to be mixed. Arentze et al. (2003) found no effects on the error variance when distinguishing between designs with two and three alternatives. However, Caussade et al. (2005) as well as Chung et al. (2011) found a U-shaped pattern with the lowest error variance for five and four alternatives respectively. A similar pattern emerged in the study conducted by DeShazo and Fermo (2002), who argued that the initial decrease of the error variance results from a better match of preferences while the increase at the later stage is caused by a more complex choice. Our hypothesis is thus that a U-shaped relationship between the number of alternatives and the error variance exists.

*Number of Attributes:* There is clear evidence suggesting that an increase in the number of attributes results in an increase in the error variance. Caussade et al. (2005) found that the number of attributes has a strong detrimental effect on a participant's ability to choose which contributes to a higher error variance. Similar inferences were drawn by DeShazo and Fermo (2002) and Arentze et al. (2003). As the information load to be processed by the respondent

grows with the number of attributes, we also expect a positive relationship between the error variance and the number of alternatives.

*Number of Attribute Levels and Level Range:* Based on the findings from Dellaert et al. (1999) and Caussade et al. (2005) and following the same arguments presented above regarding the influences on drop-out rates, we expect a positive relationship between the number of attribute levels and the error variance. Moreover, we expect an increase in the error variance with a wider level range.

### 3. Econometric approach

Random utility theory assumes the modeller does not possess complete information concerning the individual decision maker (subscript  $n$ ). Thus, individual preferences are the sum of a systematic ( $V$ ) and a random ( $\varepsilon$ ) component

$$U_{ni} = V_{ni}(x_{ni}\beta) + \varepsilon_{ni} \quad (1)$$

where  $U_{ni}$  is the true but unobservable utility associated with alternative  $i$  out of a set of available alternatives,  $V_{ni}$  is the measurable or deterministic part which itself is a function of the attributes ( $x_{ni}$ ),  $\beta$  is a vector of coefficients reflecting the desirability of the attributes, and  $\varepsilon_{ni}$  is a random term with a zero mean. This error term represents attributes and characteristics unknown to the researcher, measurement error and/or taste heterogeneity among respondents. Selection of one alternative over another implies that the utility ( $U_{ni}$ ) of that alternative is greater than the utility of the other alternative:

$$P(i) = \text{Prob}(V_i + \varepsilon_i > V_j + \varepsilon_j) \quad \forall j \in C, j \neq i \quad (2)$$

Assuming that the error components are distributed independently and identically (IID) following a type 1 extreme value distribution, one gets the multinomial logit (MNL) model where the probability of individual  $n$  choosing alternative  $i$  takes the form:

$$P_{ni} = \frac{\exp(\mu V_{ni})}{\sum_{j \in C} \exp(\mu V_{nj})} \quad (3)$$

where  $\mu$  is a scale parameter which is commonly normalised to 1 in practical applications for any one data set as it cannot be identified separately from the vector of parameters. The scale parameter is inversely proportional to the error variance  $\sigma_\varepsilon^2$ :

$$\mu = \frac{\pi}{\sqrt{6\sigma_\varepsilon^2}} \quad (4)$$

The assumption of a constant error variance across individuals has been questioned and a heteroskedastic logit model was suggested as an alternative (HL; e.g., Swait and Louviere, 1993). Here the scale parameter is no longer a constant term as it allows for unequal variances across unobserved components from two or more data sources. Whether the variation in scale across data sources can be explained by factors such as respondent characteristics or the design dimensions of the choice sets is investigated using the following HL expression (Caussade et al. 2005; DeShazo and Fermo 2002):

$$P_{ni} = \frac{\exp(\mu_n V_{ni})}{\sum_{j \in C} \exp(\mu_n V_{nj})} \quad (5)$$

where  $\mu_n = \exp(\gamma' Z_n)$  with  $Z_n$  a vector of respondent specific characteristics including the design dimensions a respondent is randomly assigned to and  $\gamma'$  a vector of parameters indicating the influence of those characteristics on the error variance (Hole 2006). The exponential form ensures a positive scale factor.

In both heteroskedastic logit models estimated in this study the scale parameter is specified as a function of the five design dimensions based on the DoD approach (number of choice sets, alternatives, attributes, attribute levels and the level ranges) as well as the respondent characteristics age, gender, and education. The two models differ with respect to the specification of the number of choice sets. In the first model (HL1) we consider this effect as linear and quadratic, while in the second model (HL2) we use dummy coded choice set

numbers ( $CS_i$ ) in order to capture scale dynamics along the sequence of choice sets where  $CS_1=0$ . The parameters  $\gamma'$  and  $\beta'$  are jointly estimated via maximum likelihood using the Stata program *clogit* (Hole 2006).

## **4. Study Design and Implementation**

### *4.1. Study Design*

Our study design largely follows the design master plan introduced by Hensher (2004). Several studies published in recent years have shown that respondents can cope with a fairly large number of choice situations. Thus, we slightly adapted the design master plan of Hensher (2004) and use 6, 12, 18 and 24 instead of 6, 9, 12 and 15 choice sets. In contrast to the Hensher (2004) study we also increased the number of attributes from four to seven instead of three to six. All other dimensions were kept equal to those of Hensher (2004). In order to obtain 16 treatments we generated 16 different generic designs using Ngene software. Unlike Hensher (2004) and Caussade et al. (2005), who employed a D-efficient experimental design, we made use of the C-error as a design criterion in each treatment as it minimises the variance of the WTP estimates (Scarpa and Rose 2008). Our adaptation of the design master plan can be seen in Table 1.

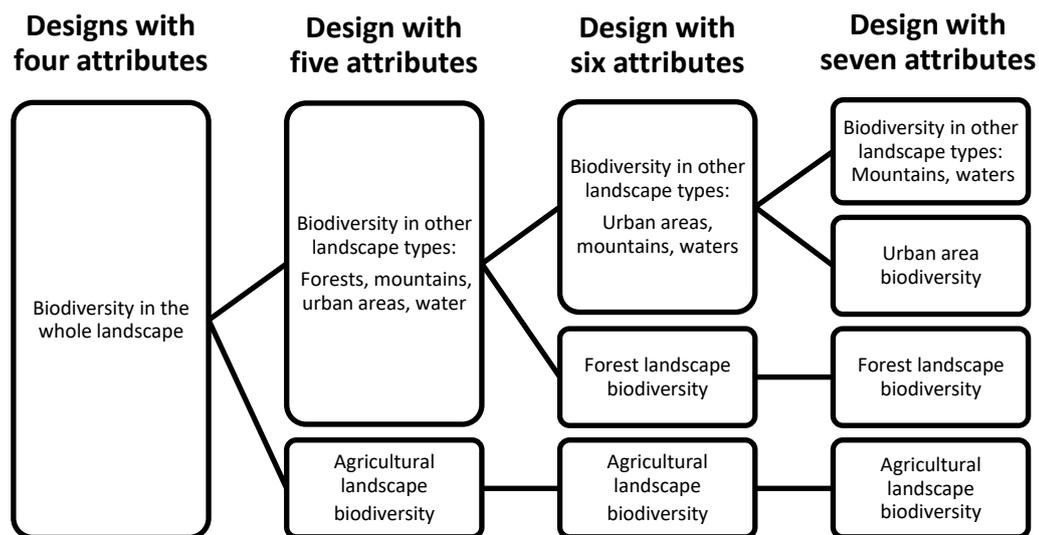
**Table 1: Adaptation of the design master plan**

<b>Design</b>	<b>Sets</b>	<b>Alternatives</b>	<b>Attributes</b>	<b>Levels</b>	<b>Range</b>
<b>1</b>	24	4	5	3	Base
<b>2</b>	18	4	5	4	+20%
<b>3</b>	24	3	6	2	+20%
<b>4</b>	12	3	6	4	Base
<b>5</b>	6	3	4	3	+20%
<b>6</b>	24	3	4	4	-20%
<b>7</b>	6	4	7	2	-20%
<b>8</b>	12	5	4	4	+20%
<b>9</b>	24	5	4	4	Base
<b>10</b>	6	5	7	3	+20%
<b>11</b>	6	4	6	4	-20%
<b>12</b>	12	5	5	2	-20%
<b>13</b>	18	4	7	2	Base
<b>14</b>	18	3	4	3	-20%
<b>15</b>	12	3	5	2	Base
<b>16</b>	18	5	6	3	-20%

The attributes in the present study deal with land use changes. For all non-price attributes the number of levels and the level range varied due to the design master plan. We distinguish between three different groups of attributes. Firstly, the attributes “share of forest” and “land conversion” were included in all treatments. The second attribute, for example, aims at the rate on converting mainly farmland into sealed land as a result of building infrastructure. The different level values were expressed in percentage changes compared to the current state. Secondly, different biodiversity attributes were used which were based on an indicator of bird populations. This indicator was developed as part of an indicator system for sustainable development in Germany (BMU 2010). The indicator can be split into bird populations in different parts of the landscape, e.g. “birds in the whole landscape” equals

“birds in agricultural landscapes” plus “birds in other landscapes”. We varied the number of attributes across treatments following the methodology used by Hensher (2004), who used different types of travel time and cost. This allowed us to aggregate and disaggregate the biodiversity attribute as a combination of already existing attributes. In doing so we were able to systematically account for the influence of the number of attributes on model outcomes (Hensher 2004). Figure 1 illustrates the split of biodiversity attributes across different designs. The levels were expressed as indicator values. Respondents were informed that an indicator of 100 or more means the landscape type is a good habitat for a variety of species.

**Figure 1: Split of the biodiversity attribute**



Thirdly, “contribution to a landscape fund” was utilised as the payment vehicle. This attribute was presented in all designs and the number of levels and level ranges were constant over all treatments. We are aware that the payment vehicle used in our study might cause problems concerning the issue of incentive compatibility. However, we decided not to use taxes due to previous experiences with protest responses in Germany. Table 2 summarises the attributes used in our study and their associated levels for a base design with two levels

and a base level range (Designs 13 and 15). The attribute levels for the other designs are reported in Appendix 1.

**Table 2: Attributes and their levels for the base design**

<b>Attribute</b>	<b>Description</b>	<b>Level</b>	<b>Unit</b>
Share of forest	Percentage changes in the share of forest	-25, +25	%
Land conversion	Percentage changes in the land conversion	-50, +50	%
Bio_whole	Biodiversity in the whole landscape including all landscape types	70, 100	indicator score
Bio_agrar	Agricultural landscape biodiversity	65, 100	indicator score
Bio_forest	Forest landscape biodiversity	80, 100	indicator score
Bio_urban	Urban area biodiversity	60, 100	indicator score
Bio_other1	Biodiversity in other landscape types: Forests, urban areas, mountains, waters	75, 100	indicator score
Bio_other2	Biodiversity in other landscape types: Urban areas, mountains, waters	60, 100	indicator score
Bio_other3	Biodiversity in other landscape types: Mountains, waters	75, 100	indicator score
Cost	Contribution to a landscape fund per year	10, 25, 50, 80, 110, 150	€

As it can be seen from Table 1, the number of alternatives varies from three to five including the status quo alternative, which was defined as the current situation with the cost attribute of zero. The attribute levels for the status quo alternative were “as today” as we asked respondents to assess land use changes for an area within 15 km of their residence without knowing the shape of the landscape in each respondent’s surrounding. This is another factor that distinguishes our study from Hensher (2004). Further differences are summarised in Table 3.

**Table 3: Differences across complexity studies**

<b>Characteristic</b>	<b>Hensher (2004)*</b>	<b>Present Study</b>
Application	Transportation	Environmental valuation
Number of choice sets	6, 9, 12, 15	6, 12, 18, 24
Number of attributes	3 – 6	4 – 7
Survey mode	Computer aided personal interview (CAPI)	Online survey
Experimental design	D-efficient	C-efficient
Status quo	Current route	Current situation
Region	Sydney	Germany (nationwide)
Aim of assessment	Route changes	Land use changes around 15 km of residence
Variation of attribute numbers	Based on different types of travel time and of costs	Based on Biodiversity in different landscapes types
Payment vehicle	Travel cost	Contribution to a landscape fund
Number of observations	8,020	90,354

\* see also Caussade et al. (2005), Rose et al. (2009), Chintakayala et al. (2010)

#### 4.2. *Survey Implementation*

A nation-wide online survey was conducted between December 7th and December 21st 2012 using a panel from a survey company. When participants entered the survey, they were randomly assigned to one of the 16 designs. The questionnaire started by asking respondents socio-demographic questions concerning date of birth, gender and education. We did this at the initial stage of the questionnaire in order to be able to include socio-demographic characteristics in the drop-out analysis. Then, participants were asked several ‘warm-up’ questions in order to introduce them to the topic and to obtain their assessment of the perceived current state of landscape characteristics within 15 km of their residence. Before presenting the first choice set, participants were given an instruction page with information on the CE as well as the attribute descriptions. These descriptions varied across

treatments with four, five, six and seven attributes. In order for the choice sequence effect not to be confounded with the choice task effect, we randomised the sequence of choice tasks within each design. So, the position of each choice set differed across respondents. For each choice set respondents were asked to assess land use changes for an area within 15 km of their residence by choosing their preferred alternative. Figure 2 shows an example choice set. For the biodiversity attributes the value of today's indicator score was presented on each choice card. Additionally, respondents were informed that these values could vary across regions in Germany. For the attributes "share of forest" and "land conversion" we do not know the present situation in the 15km surrounding of each respondents. Thus we did not provide any values for the current situation and used relative changes as attribute levels.

**Figure 2: Example choice set with four alternatives and seven attributes**

<p>If only the following options were available for the future development of the landscape within a <u>radius of up to 15 kilometres around your place of residence</u>, which one would you choose?</p> <p><i>If you live in a large city, please consider the surrounding area of the city.</i></p>				
	<b>Option 1</b>	<b>Option 2</b>	<b>Option 3</b>	<b>Option 4</b>
<b>Share of forest in the landscape</b>	20% more	20% less	20% more	as today
<b>Land conversion</b>	40% more	40% more	40% more	as today
<b>Biodiversity in agricultural landscapes</b> (Today 65 scores)	76 scores	89 scores	89 scores	as today
<b>Biodiversity in forest landscapes</b> (Today 80 scores)	95 scores	95 scores	95 scores	as today
<b>Biodiversity in other landscape types</b> (Today 65 scores)	89 scores	89 scores	89 scores	as today
<b>Biodiversity in urban areas</b> (Today 60 scores)	90 scores	90 scores	70 scores	as today
<b>Financial contribution to the landscape fund per year</b>	80 €	50 €	80 €	0 €
<b>I choose</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

After finishing the CE several follow-up questions were asked including perceived choice certainty and subjective attendance to attributes. As respondents progressed through the questionnaire, they saw a bar that indicated what fraction of the questionnaire had been answered. According to the number of choice sets respondents saw slower progression of the bar when they were assigned to a design with more choice sets. However, respondents were not informed in advance about how many choice sets they would face during the interview.

At each stage of the questionnaire respondents could only abandon the survey by closing their web browser. In this case the drop-out position was recorded. In total, 2122 interviews were collected with 1684 (79.36%) participants completing the whole questionnaire and 438 participants only partially completing it. For the group of respondents who quitted before the CE we do not have sufficient data as many respondents dropped out on the initial screens before any socio-demographic characteristics were requested. Therefore, we do not include this group in subsequent analysis. The average interview length was measured to be 23 minutes and the response rate was 29.49% (corresponding to the completed interviews relative to persons contacted). Table 4 reports basic socio-demographic characteristics for the sample including those respondents who started to answer the CE (completed plus drop-outs) as well as the sample used to estimate the choice models (completed interviews). Education refers to the highest educational level achieved. Across the 16 designs we could not find any statistically significant differences regarding the socio-demographic characteristics for both samples reported in Table 4.

**Table 4: Socio-demographic characteristics**

Variable		Sample including drop outs* mean (SD)	Sample completed interviews mean (SD)
Age in years		42.75 (13.69)	42.34 (13.59)
Gender (male = 1)		0.50 (0.50)	0.48 (0.50)
Highest educational level	Compulsory secondary school (9 years schooling)	0.07 (0.24)	0.07 (0.24)
	Middle-school degree (10 schooling years)	0.28 (0.45)	0.27 (0.44)
	High-school degree (13 schooling years)	0.26 (0.44)	0.27 (0.44)
	University degree	0.39 (0.49)	0.39 (0.49)
Number of respondents		1,947	1,684

Note: Middle-school is used here as an equivalent of the German “Realschule”. \*This sample does not comprise those 175 respondents who left the survey before the CE started, SD =standard deviation

## 5. Results

### 5.1. *The Influence of the Design Dimensionality on Drop-out Rates*

As a first empirical illustration, Table 5 depicts the drop-out rates according to the stage of completion of the questionnaire. Around 8.2% of the participants dropped out before starting to answer the CE. The highest drop-out rate is within the CE (10.8%), while only 1.6% quit the survey after finishing the CE. The relatively high drop-out rate in the first part of the questionnaire has also been observed in studies carried out in other areas of research. Hoerger (2010) also found 10.0% of the participants to drop out of the survey instantaneously. This behaviour may be explained by a low interest in the topic. Among others, Galesic (2006) found that the lower the respondent’s interest in the topic of the survey, the higher the drop-out rate. The authors asked respondents at the end of each of the 20 question blocks, which contained questions on a specific topic, how interesting they

found the questions in the preceding block. For respondents who dropped out at any later point in the survey levels of interest were lower at previous blocks.

**Table 5: Position of drop-outs**

<b>Drop-out position</b>	<b>Frequency</b>	<b>%</b>
Before CE	175	8.25
Within CE	229	10.79
After CE	34	1.60
Completed	1684	79.36
Total	2122	100.00

Table 6 presents the 16 treatments with their corresponding design dimensions as well as the number of interviews and the number of drop-outs per design. We exclude those respondents who quit the survey before starting to answer the CE as all questions were identical across designs up to this stage. In total 263 respondents dropped out of the survey within and after the choice task section of the interview. The designs with the highest drop-out rates are the designs 9, 3 and 1. Each has 24 choice sets with all other design dimensions varying (three to five alternatives, four to six attributes, etc.). The three designs with the next highest drop-out rates all have 18 choice sets. At the other end of the spectrum we have the designs 2, 15, 5 with the lowest drop-out rates. All these designs have a different number of choice sets, but none has 24 choice sets. Thus, there is already some indication that drop-out rates are associated with the number of choice sets.

**Table 6: Design-dependent drop-outs within and after the choice experiment**

Design	Drop-outs			Design Dimensions				
	In %	Number	Complete interviews	Sets	Alternatives	Attributes	Levels	Range
9	25.93	21	81	24	5	4	4	Base
3	25.00	31	124	24	3	6	2	20%
1	21.18	18	85	24	4	5	3	Base
13	20.35	23	113	18	4	7	2	Base
14	20.24	17	84	18	3	4	3	-20%
16	20.24	17	84	18	5	6	3	-20%
10	18.54	28	151	6	5	7	3	20%
8	16.46	13	79	12	5	4	4	20%
6	15.85	13	82	24	3	4	4	-20%
4	15.00	12	80	12	3	6	4	Base
12	12.35	10	81	12	5	5	2	-20%
11	11.36	10	88	6	4	6	4	-20%
7	11.26	25	222	6	4	7	2	-20%
2	10.98	9	82	18	4	5	4	20%
15	6.94	10	144	12	3	5	2	Base
5	5.77	6	104	6	3	4	3	20%
<b>Total</b>	15.62	263	1684					

In order to analyse the relationship between the five design dimensions and drop-out rates in detail, we specify a binary logit model with the dependent variable being zero if a participant completed the survey and one if the respondent dropped out after starting to answer the CE. The results are presented in Table 7. As expected the number of choice sets has a highly significant positive impact on the probability to drop out. The same pattern is observed for the number of attributes. Regarding the number of alternatives, only the dummy variable for five alternatives significantly influences the probability to drop out. This effect is, however, only significant at the 10% level. For the attribute levels and the level range we find no significant effect. So, we could not reject our null hypothesis of no effect. A linear regression on the influence of the design dimensions on the number of respondents who dropped of the survey confirms these findings. The results also show that the number of choice sets as well as the

number of attributes positively influence the number of respondents who drop out (see Appendix 2 for the linear regression results).

Since we also expected socio-demographic characteristics to be possible explanations for the probability to drop out, we included age, gender and education as further independent variables. As shown in Table 7 all but the second age group have a significantly higher probability to drop out compared to respondents being between 18 and 30 years old. Also the dummy for being male is positive and highly significant. With respect to education all parameters have a negative sign indicating that higher education leads to less drop-outs compared to the reference group. However, only the parameter for high-school degree is statistically significant. Finally, none of the interactions among design dimensions we incorporated in the model are statistically significant at the 5% level.

**Table 7: Results of the binary logit model on drop-outs**

<b>Dimension for the probability to drop out</b>	<b>Coefficient</b>	<b>( t-Value )</b>
Number of choice sets	0.0556	4.80
Dummy 4 alternatives	-0.0428	0.22
Dummy 5 alternatives	0.3355	1.85
Number of attributes	0.2060	2.32
Narrow level range	0.0727	0.42
Wide level range	0.0856	0.46
Number of levels	0.0720	0.68
Age group 31 to 40 years	0.4075	1.86
Age group 41 to 50 years	0.1094	0.51
Age group 51 to 60 years	0.5079	2.33
Age group older than 60 years	0.4906	2.09
Gender male	0.6888	4.90
Middle-school degree	-0.1680	0.65
High-school degree	-0.6169	2.19
University degree	-0.1994	1.44
Constant	-4.5138	5.46

Log-likelihood null	-770.88
Log-likelihood model	-734.69
Pseudo-R <sup>2</sup>	0.05

Note: N = 1,947 including 263 respondents who dropped out within or after the CE

### 5.2. Drop-outs within the Choice Experiment

To conclude our analysis on drop-out rates, Table 8 shows the drop-out position within the CE. We observe the highest number of participants (123) to drop out within the first six choice sets. This figure corresponds to 53.7% of the drop-outs within the CE and a drop-out rate of 6.3%. At the other extreme, only 20 respondents or 8.7% of the drop-outs abandoned the survey between choice set 19 and set 24. The total drop-out rate at this stage of the CE is therefore only 4.4%. Therefore, respondents who do not agree with the format of a choice experiment seem to drop out early while those who agree are likely to stay even if they face a longer sequence of sets.

**Table 8: Drop-out rates within the choice experiment**

Position drop out	Number of presented choice sets				Total	Respondents in respective designs	Drop-out rate
	6 Sets	12 Sets	18 Sets	24 Sets			
Set 1 to 6	57	20	21	25	123	1947	6.3
	<i>24.9</i>	<i>8.7</i>	<i>9.2</i>	<i>10.9</i>	<i>53.7</i>		
Set 7 to 12		16	19	16	51	1191	4.3
		<i>7.0</i>	<i>8.3</i>	<i>7.0</i>	<i>22.3</i>		
Set 13 to 18			16	19	35	849	4.1
			<i>7.0</i>	<i>8.3</i>	<i>15.3</i>		
Set 19 to 24				20	20	454	4.4
				<i>8.7</i>	<i>8.7</i>		
	57	36	56	80	229		
	<i>24.9</i>	<i>15.7</i>	<i>24.5</i>	<i>34.9</i>	<i>100.0</i>		

Note: percentages in italics

### 5.3. Heteroskedastic Logit Models

We estimated three different models to investigate the association between the design dimensions and scale. First, we report the results from a simple MNL model in which the data

from all split samples are pooled and only the choice attributes are incorporated. Next, estimates from a heteroskedastic logit model (HL1) with the scale parameter as a function of the design dimensions, interactions among them, and socio-demographic characteristics are reported in Table 9. Another heteroskedastic logit model (HL2) investigating order effects along the sequence of choice sets is presented in Appendix 2.

In all three models the attribute parameters are deemed significant at a 1% level of significance and have the expected sign. On average respondents prefer a larger share of forests while at the same time preferring reduced land conversion in their surroundings. The biodiversity attributes all have positive signs indicating that respondents want to increase the levels of biodiversity as measured by the underlying scale. The model also contains an alternative specific constant (ASC<sub>sq</sub>) for the current situation. It is positively and statistically significant suggesting that, on average, respondents have a propensity to choose the current situation instead of one of the hypothetical alternatives describing future land use changes.

**Table 9: Estimation results for the MNL and HL1 model**

Variable	MNL		HL1	
	Coefficient	t-Value	Coefficient	t-Value
ASCsq	0.5114	19.04	0.2230	2.99
Share of forest	0.0167	35.57	0.0079	3.02
Land conversion	-0.0085	38.31	-0.0041	3.03
Bio_whole	0.0096	8.61	0.0053	2.95
Bio_agrar	0.0054	9.36	0.0025	2.81
Bio_forest	0.0059	6.75	0.0029	2.70
Bio_urban	0.0061	5.81	0.0032	2.63
Bio_other1	0.0076	6.73	0.0033	2.76
Bio_other2	0.0027	3.85	0.0012	2.18
Bio_other3	0.0052	5.81	0.0022	2.52
Cost	-0.0062	26.51	-0.0028	3.06
<i>Covariates</i>				
Position of choice set			0.0354	4.16
Squared position of choice set			-0.0008	2.45
Dummy 4 alternatives			0.2262	3.35
Dummy 5 alternatives			0.0731	1.16
Interaction 4 alternatives * number CE			-0,0139	2,57
Interaction 5 alternatives * number CE			-0,0027	0,50
Number of attributes			0.1046	1.78
Number of levels			0.3147	3.15
Interaction attributes*levels			-0.0658	3.44
Narrow level range			0.2046	5.54
Wide level range			-0.1042	2.69
Age31to40			-0.0121	0.31
Age41to50			-0.0812	2.16
Age51to60			0.0587	1.50
Age60plus			-0.1672	3.54
Gender male			-0.1279	4.95
Middle-school degree			-0.0018	0.03
High-school degree			0.0954	1.61
University degree			0.1449	2.57
Log-likelihood null		-31114.18		
Log-likelihood model		-27438.90		-27291.44
Observations		90354		

The estimates from the HL1 model regarding the effect of the design dimensions, interactions among them, and the socio-demographics characteristics on scale are reported in the lower part of Table 9. This model accounts for the number of choice sets by incorporating the position of each choice set as a linear and quadratic effect. Note that the

choice sets were always presented in a randomized order for each design. Similar to Caussade et al. (2005) we find evidence for an inverse U-shaped relationship between the number of choice sets and scale. The linear effect shows that an increasing number of choice sets affects scale positively indicating a learning effect. As the respondent becomes familiar with the choice tasks and moves along the sequence of choices, a decrease in variance (and hence an increase in the scale of the Gumbel error) is observed. This has been attributed to learning. At the same time, though, an opposite effect is observed that has been attributed to increase in respondent's fatigue and results in an increased error variance (and hence a lowered Gumbel error scale). Towards the end of the choice sequence the latter effect overrides the former, thereby producing an overall increase in error variance.

With respect to the number of alternatives, we can reject the null hypotheses by observing a U-shaped pattern with dummy variables specified for four and five alternatives. Designs with four alternatives have a higher scale parameter (lower error variance) when compared to the designs with three alternatives. This suggests a similar pattern as observed by Caussade et al. (2005). However, the coefficient for the variable indicating five alternatives per choice set is not statistically significant in our model. Regarding both the number of attributes and the number of levels we find significant impacts on the scale parameter, but the effect of the number of attributes is only significant at the 10% level. Our data also rejects the null hypothesis of no effect on scale due to the last of our design dimensions: the range of attribute levels. We observe the error variance to decrease for designs with a narrow range and to increase for designs with a wide range compared to the base. Choice consistency is promoted when level ranges are narrow as comparisons among alternatives are easier to carry out. This result is again in line with Caussade et al. (2005).

Testing whether the effect of one design dimension depends on another we found two statistically significant interactions. Firstly, the interaction between the choice set position and the number of alternatives reveals that scale increases until the twentieth choice set.

Afterwards it remains at the same level (Figure 3). For choice sets with four alternatives, however, we find that after a significant increase in scale at the beginning of the sequence and a modest increase between the choice sets 2 and 12, scale decreases resulting in an increased error variance in the second half of the sequence. The interaction for five alternatives and the position of the choice set is, however, not statistically significant. Thus, designs with three or five alternatives do not cause different error variances. Secondly, the highly significant interaction effect between number of attributes and number of levels indicates that the impact of both design dimensions on scale is conditional. An increasing number of attributes or attribute levels, everything else remaining equal, decreases the scale effect.

We find significant effects for all of the following socio-demographic covariates: age, gender, and education. Two of the age groups have a statistically significant lower scale parameter in contrast to the reference group of respondents age 18 to 30. Also, being male has a negative influence on the scale parameter while holding a university degree has a positive influence on scale compared to those with nine and less years schooling (the reference group). To compare both the MNL and HL1 model we use a log-likelihood ratio test. The test statistic is 147.46. Since the critical Chi-squared value is 33.41 (1% level with 17 degrees of freedom), we can reject the null hypothesis that the HL1 model is no better than the MNL model.

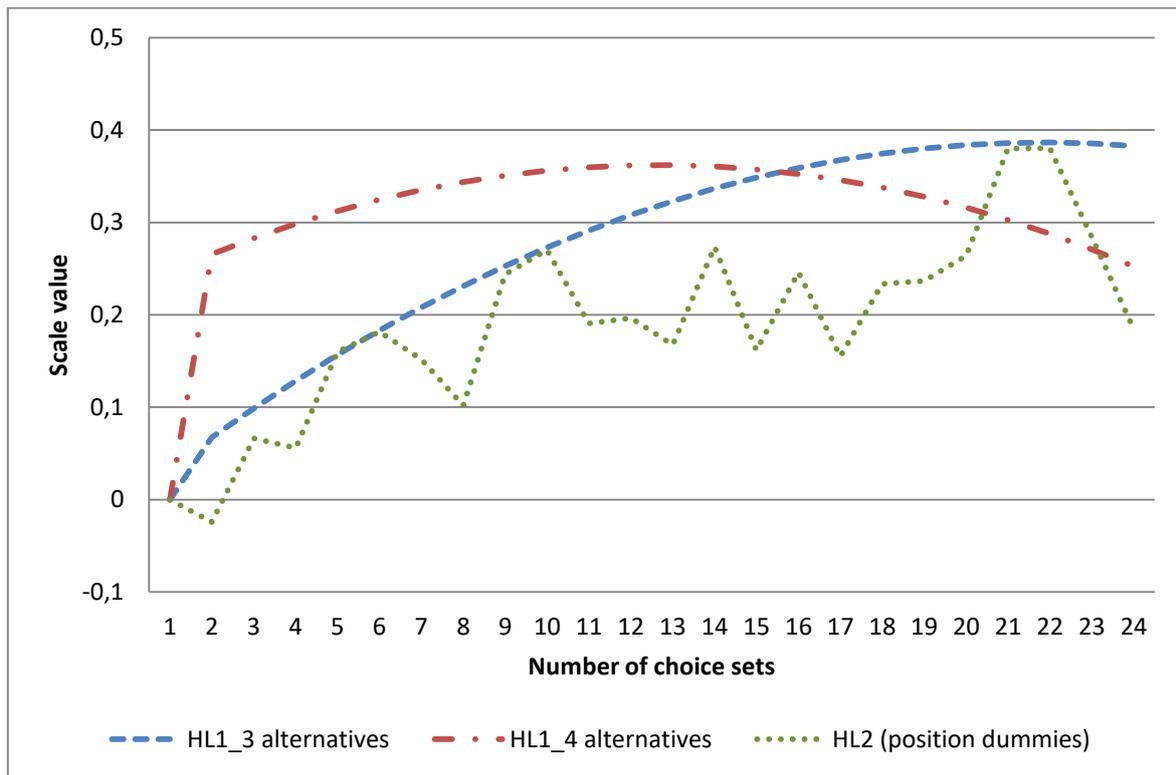
We now turn to the HL2 model accounting for order effects with dummies for each choice set except for the first ( $CS_2$ - $CS_{24}$ ; Appendix 2). Note that this model does not, unlike model HL1, simultaneously account for the number of alternatives per choice set due to the very high number of additional interactions required by such a model. The results suggest an order effect as the  $CS_2$ - $CS_{24}$  coefficients are jointly significant. In contrast to other studies (Scarpa et al. 2011; Czajkowski et al. 2012; Carlsson et al. 2012), however, we do not find a clear learning effect on the first choice sets. The coefficients for the choice sets  $CS_2$  to  $CS_4$  are not statistically significant, the coefficient for  $CS_3$  has a negative sign. The first dummy variable

positively significant is related to  $CS_5$ . The highest values occur, on the other hand, for the sets  $CS_{21}$  and  $CS_{22}$ , before the value decreases again. As the coefficient for the last choice set is not statistically significant, it might not be justified to argue that this drop in scale indicates fatigue.

The scale effects for both models are illustrated in Figure 3. Firstly, the dashed line (colored red) shows the scale effects due to model HL1 for the reference alternative with three alternatives per choice set. This line indicates a low but constantly rising scale value until  $CS_{21}$ - $CS_{22}$  which afterwards remains at that level. Secondly, the line with dashes and dots (colored red) shows the scale effect for choice sets with four alternatives. For these designs we find a significantly positive effect at the beginning of the sequence. However, this effect decreases more and more when we move along the sequence of sets. After  $CS_{16}$  scale becomes lower than in the reference situation with three alternatives. This suggests that for designs with more than 16 sets it is advantageous to use three or five alternatives. Also, the latter does not show statistically significant effects on scale compared to designs with three alternatives. Finally, the dotted line (colored green) shows the scale effects for each choice set based on model HL2 (see Appendix 2). As these values capture the effect across designs with three, four and five alternatives, the values are always below the dashed blue line.

Overall, the results from both the HL1 and the HL2 models seem to support findings presented by Czajkowski et al. (2012; similar Hess et al. 2012). They found that scale rather remained on the same level after it has been increasing for the first eight to nine choice sets, thus, not clearly suggesting fatigue effects at the end of the sequence of choices. For designs with four alternatives, however, our results indicate fatigue suggesting that this effect differs across designs with a varying number of alternatives.

**Figure 3: Scale effects along the order of choice sets**



Note: In HL2 the dummies for the positions 2, 3, 4, 8, and 24 are not statistically significant at the 10% level

Table 10 reports the marginal WTP estimates for the MNL and the HL1 model. The confidence intervals were calculated using the Delta method. Note that the marginal WTP estimates refer to environmental changes in the area within 15 km of each respondent's residence. Starting with the estimates based on the MNL model, respondents are, on average, willing to pay 2.71 € per year for a one percent increase in the share of forests, but would experience a disutility of 1.38 € per year for a one percent growth in land conversion. The biodiversity attributes, which were aggregated and disaggregated across designs, delivered results as expected. Respondents are willing to pay more for the aggregated attribute "whole landscape biodiversity" (Bio\_whole), which is 1.55 € per year for an increase in the indicator score, than for attributes at a lower aggregation level such a forest landscape biodiversity (0.95 € per year for an improvement of a score) or agricultural landscape

biodiversity (0.88 € per year for an improvement of a score). Moreover, the WTP for Bio\_other1, which includes biodiversity in forests, urban areas, mountains and waters, is higher than Bio\_other2 considering biodiversity in mountains, urban areas and waters. In contrast, the WTP for Bio\_other3 is close to the WTP value for Bio\_other1 although it comprises less habitats.

**Table 10: Marginal willingness to pay estimates for attributes in Euro per year**

Attribute	MNL		HL1		Diff MNL – HL
	mWTP €	95%-CI	mWTP €	95%-CI	One-sided significance mWTP <sub>MNL</sub> - mWTP <sub>HL1</sub>
Share of forest	2.71	2.43 / 2.99	2.85	2.55 / 3.15	0.38
Land conversion	-1.38	-1.50 / -1.27	-1.47	-1.60 / -1.34	0.33
Bio_whole	1.55	1.17 / 1.94	1.90	1.49 / 2.30	0.15
Bio_agrar	0.88	0.68 / 1.09	0.91	0.68 / 1.14	0.49
Bio_forest	0.95	0.66 / 1.25	1.04	0.69 / 1.39	0.49
Bio_urban	0.98	0.64 / 1.32	1.15	0.74 / 1.56	0.28
Bio_other1	1.24	0.86 / 1.62	1.19	0.79 / 1.60	0.42
Bio_other2	0.44	0.21 / 0.67	0.43	0.16 / 0.70	0.44
Bio_other3	0.86	0.56 / 1.15	0.79	0.44 / 1.15	0.37

Note: The one-sided p-values for the difference between both WTP estimates result from 2500 bootstrap replications; CI =confidence interval.

The absolute WTP estimates differ between the MNL and the HL1 models, but the point estimates are each time included in the interval of the other model for the same attribute. This indicates that the WTP estimates are not statistically different. Additionally, we directly bootstrapped the difference between both WTP estimates for each attribute using 2500 repetitions. In order to determine the approximate one-sided significance of the difference we follow the percentile approach and calculate the proportion of negative values of the difference between the marginal WTP estimates from both models (see Poe et al. 1997). The calculated values (last column Table 10), resulting from 2500 replications, support the conclusion that the difference between both models is not statistically significant. Thus,

taking into account the impact of the design dimension and socio-demographic characteristics on the error variance does not result in significantly different WTP estimates.

## **6. Discussion**

In this study we have analysed the impact of the number of choice sets, the number of alternatives in each choice set, the number of attributes as well as the associated number and range of levels on drop-out rates and the error variance. To our knowledge it is not only the first study employing the Design of Designs approach introduced by Hensher (2004) in environmental valuation, but also the first attempt to investigate the relationship between task complexity in discrete CE and participants' drop-outs in this field.

With respect to the drop-out rates, we find that the probability to abandon the survey significantly increases with the number of choice sets and the number of attributes. Furthermore, designs with five alternatives show a higher drop-out rate compared to those with only three alternatives. All other design dimensions do not significantly influence the probability to quit the survey. Additionally, older respondents as well as men are more likely to drop out. Among those who dropped out while answering the choice tasks, it is noteworthy that the majority of the respondents abandoned the CE within the first six choice questions. One reason for this could be that people did not like the choice format and thus decided not to proceed. During focus groups we have repeatedly discovered that some respondents do not like to make comparisons among bundles of attributes, but would prefer to rate each attribute separately. Another reason might be that people realised that choosing among the alternatives on a choice set also affects payments to the landscape fund and that quitting the survey at this stage was motivated by protest votes. However, we do not conclusively know the reasons why respondents abandoned the survey; we refrained from sending those respondents debriefing questions as the survey company expected very low response rates for this kind of questions.

The results overall suggest that if only drop-out rates in an online survey are of concern, fewer choice sets and attributes as well as less than five alternatives should be presented. However, Louviere et al. (2013) argue that due to the cost of statistical information, i.e. the survey costs per interview, the potentially negative effect of designs with a higher dimensionality on drop-out rates is overcompensated by the additional information gained. Applying their calculations (see Louviere et al. 2013: 28) to our data we can confirm this finding. The design with the lowest drop-out rate (Design 5: 6 choice sets, 3 alternatives, 4 attributes) would result in 1131 data records per 100 individuals starting the survey (and 94 finishing it). The design with the highest drop-out rate (Design 9: 24 choice sets, 5 alternatives, 4 attributes) would result in 711 data records per 100 individuals starting to answer the survey (and 74 finishing it). Clearly, the design with a higher dimensionality would provide more information although the completion rate is significantly lower. Generally, drop-out rates and the reasons why respondents drop out of a survey have not found much attention. This is particularly surprising for online surveys taking into account how easily this can be investigated using paradata. Whether drop-out rates depend on survey mode is another interesting question. In a computer assisted personal interview (CAPI), as used by Czajkowski et al. (2012), for example, respondents might on average be more motivated or might feel more obliged to go through a longer sequence of choice sets.

With respect to influences of the design dimensions on the error variance, we mainly find the same results as Caussade et al. (2005). For all five dimensions we reject our null hypothesis of no relationship between the design dimensionality and the error variance. However, there are some differences compared to the findings by Caussade et al (2005) in transportation. The first is that fatigue effects are not as clearly present as in their study. For designs with three and five alternatives scale does not decrease as rapidly as in their study or as presented by other authors (e.g., Scarpa et al. 2011). In contrast, for designs with four alternatives we firstly observe a significant increase followed by a decreasing scale value. An explanation for this could be that there is an easier choice on sets with an odd number of

alternatives. Overall, our findings are rather in line with results presented by Czajkowski et al. (2012) and Hess et al. (2012) who could not find strong evidence for fatigue effects due to the number of choice sets. The second key difference is that overall no large increase in choice certainty is observable at the first choice sets in the sequence; only for designs with four alternatives we notice a stronger increase. This raises the question whether the first choices should be used as a valid expression of preferences or whether they should be seen solely as warm-up exercises. That choice certainty does not increase strongly at the beginning also raises the question whether instruction choice sets, as suggested for example by Carlsson et al. (2012), would counteract this successfully. The third difference worth emphasizing is that in our study also two interactions among design dimensions are statistically significant while Caussade et al. (2005) could not discover any significant interaction. As already mentioned, the position of the choice set in the sequence of choices and the number of alternatives interact. Designs with four alternatives perform better at the beginning of the sequence but within the course of that sequence the scale value significantly decreases indicating that this format is less suitable for longer sequences of choice sets. The second highly significant interaction effect is between the number of attributes and the number of attribute levels. A low number of attributes or levels has a positive effect on scale that decreases and turns into a negative effect when the number of attribute levels, all else remaining equal, increases.

For all attributes considered in our study the WTP estimates are calculated based on the MNL and the HL1 model. The WTP estimates are higher for an aggregated biodiversity attribute than for biodiversity attributes at a lower level of aggregation, except in one case where the WTP estimate for Bio\_oth3 is higher than for Bio\_oth2. This pattern points to valid WTP estimates. Comparing the estimates from the MNL and the HL1 model does not reveal significantly different values. Accounting for the impact of the dimensionality on scale differences does therefore not change the main findings regarding the WTP measures. This finding is again in line with the results presented by Caussade et al. (2005).

Overall, our results confirm those findings presented in the literature that indicate that a higher dimensionality does result in more information gained even when drop-out rates increase. The WTP estimates that are not statistically different across models indicate that this is not the case. However, our results should be interpreted with some degree of caution. Firstly, our payment vehicle, a contribution to a fund, might not be as incentive compatible as the costs respondents face in transportation surveys dealing with a private good. Secondly, following largely the modelling approach presented by Caussade et al. (2005) we have not addressed taste heterogeneity so far. As the survey was conducted nation-wide and respondents lived in quite different landscapes, it is likely that participants prefer different changes in their landscapes. Accounting for both observed and unobserved heterogeneity thus is likely to have an impact on the results. Studies investigating the relationship between taste and scale heterogeneity found that both are associated with each other (Dellaert et al. 2012; Czajkowski et al. 2012; Scarpa et al. 2011).

In future analysis more flexible models in terms of taste heterogeneity will be applied. Also, we will investigate to which extent the design dimensions influence the number of times the status quo alternative is chosen. The results presented so far in the literature indicate that the status quo option is more likely to be chosen when the choice design is more complex (Boxall et al. 2009; Zhang and Adamowicz 2011). Additionally, effects of the choice task on choices and WTP estimates will be analysed using various measure of complexity (Dellaert et al. 2012; DeShazo and Fermo 2002; Swait and Adamowicz 2001). This could provide further insights about the impact of the design dimensions on the results of stated CE suggesting whether researchers need to care more about this issue, at least within the ranges investigated in this study. Another important topic is, particularly if a larger number of choice sets is used to gain more information per respondent, whether sequences of choice questions are incentive compatibility. Vossler et al. (2012) show that under certain conditions sequences of binary choice questions are incentive compatible. Whether this finding is valid

also for other choice question formats, e.g. with three and more alternatives, or other survey contexts is so far an open question.

### **Acknowledgement**

We would like to thank two anonymous reviewers for their very valuable comments. Also we would like to thank Helen Lauff (Link-Institut) for her advice and patience while implementing our Design of Designs approach in the survey software. Funding for this research was provided by the Federal German Ministry of Education and Research (Fkz. 033L029G; Fkz 01LL0909A) and is gratefully acknowledged .

### **References**

- Arentze T, Borgers A , Timmermans H, DelMistro R (2003) Transport Stated Choice Responses: Effects of Task Complexity, Presentation Format and Literacy. *Transportation Research Part E* 39: 229–244.
- Bech M, Kjaer T, Lauridsen J (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics* 28(3): 273-286.
- Boxall P, Adamowicz W L , Moon A (2009) Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Aust J Agr Resour Economics* 53: 503-519.
- Bradley M, Daly A (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation* 21: 167-184.
- Brouwer R. Dekker T, Rolfe J, Windle J (2010) Choice Certainty and Consistency in Repeated Choice Experiments. *Environ Resource Econ* 46: 93-109.
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.

- Carlssona F, Mørkbak M R, Olsen S B (2012) The first time is the hardest: A test of ordering effects in choice experiments. *Journal of Choice of Modelling* 5: 19-37.
- Caussade S, Ortúzar J, Rizzi L I, Hensher D A (2005) Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, 39: 621-640.
- Chintakayala P K, Hess S, Rose J M, Wardman M (2010) Effects of Stated Choice Design Dimensions on Model Estimates. In: Hess S, Daly A (ed) *Choice Modelling: State-of-the-Art and the State-of Practice*. Proceedings from the Inaugural International Choice Modelling Conference, Emerald Press, Bingley.
- Czajkowski M, Giergiczny M, Greene W H (2012) Learning and Fatigue Effects Revisited. The Impact of Accounting for Unobservable Preference and Scale Heterogeneity on Perceived Ordering Effects in Multiple Choice Task Discrete Choice Experiments. forthcoming in *Land Economics*.
- Chung C, Boyer T, Han S (2011) How many choice sets and alternatives are optimal? Consistency in choice experiments. *Agribusiness* 27: 114-125.
- Dellaert B, Brazell J, Louviere J J (1999) The effect of attribute variation on consumer choice consistency. *Marketing Letters* 10: 139-147.
- Dellaert B, Donkers B, van Soest A (2012) Complexity Effects in Choice Experiment-Based Models. *Journal of Marketing Research* 49: 424-434.
- DeShazo J R, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123-143.
- Galesic M. (2006) Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics* 22: 313-328.
- Heiner R (1983) The origin of predictable behaviour. *The American Economic Review* 73: 560-595.

- Hensher D A, Stopher P A, Louviere J J (2001) An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application. *Journal of Air Transport Management* 7: 373-379.
- Hensher D A (2004) Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. *Journal of Transport Economics and Policy* 38: 425-446.
- Hensher D A (2006) Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment. *Environ Resource Econ* 34: 7-44.
- Hess S, Hensher D A, Daly A (2012) Not bored yet – Revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A*: 626–644.
- Hoerger M (2010) Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. *Cyberpsychology, Behavior and Social Networking* 13: 697-700.
- Hole A R (2006) Small-sample properties of tests for heteroscedasticity in the conditional logit model. *Economics Bulletin* 3: 1-14.
- Louviere, J. J., Carson, R. T., Burgess, I., Street, D., Marley, A.A.J. (2013) Sequential preference questions factors influencing completion rates and response time using online panel. *The Journal of Choice Modelling* 8: 19-31
- Louviere J J, Islam T, Wasi N, Street D, Burgess, L.B. (2008) Designing Discrete Choice Experiments: Do Optimal Designs Come At A Price? *Journal of Consumer Research* 35: 360-375
- Louviere, J.J., D.A., Hensher, Swait J.D. (2000) *Stated Choice Methods. Analysis and Application*. Cambridge.
- McNair B J, Bennett J, Hensher D A (2011) A comparison of responses to single and repeated discrete choice questions. *Resource and Energy Economics* 33: 554-571.

- Poe G L, Welsh M P, Champ P A (1997) Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics* 73: 255-267.
- Ratcliffe J, Longworth L (2002) Investigating the structural reliability of a discrete choice experiment within health technology assessment. *International Journal of Technology Assessment in Health Care* 18: 139-144.
- Rolfe J, Bennett J (2009) The impact of offering two versus three alternatives in choice modelling experiments. *Ecological Economics* 68: 1140-1148.
- Rose J M , Hensher D A, Caussade S, Ortúzar J, Jou R-C (2009) Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. *Journal of Transport Geography* 17: 21-29.
- Ryan M, Wordsworth S (2000) Sensitivity of Willingness to Pay Estimates to the Level of Attributes in Discrete Choice Experiments. *Scottish Journal of Political Economy* 47: 504-524.
- Scarpa R, Notaro S, Louviere J, Raffaelli R (2011) Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in Alpine grazing commons. *American Journal of Agricultural Economics* 93: 813-828.
- Scarpa R, Rose J M (2008) Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *The Australian Journal of Agricultural and Resource Economics* 52: 253-282.
- Swait J D, Adamowicz W L (2001) Choice environment, market complexity and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behaviour and Human Decision Processes*, 86: 141-167.
- Vicente P, Reis E (2010) Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. *Social Science Computer Review* 28: 251-267.

Vossler C A, Doyon M, Rondeau D (2012) Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. *American Economic Journal: Microeconomics* 4: 145-171.

Zhang J, Adamowicz W L (2011) Unravelling the Choice Format Effect: A Context-Dependent Random Utility Model, *Land Economics* 87: 730-743.

## Appendix 1

**Table 1: Attribute levels for designs with two levels**

Attribute	narrow		base		wide	
Share of forest (%)	-20	+20	-25	+25	-30	+30
Land conversion (%)	-40	+40	-50	+50	-60	+60
Bio_whole	80	90	70	100	53	117
Bio_agrar	76	89	65	100	49	116
Bio_forest	85	95	80	100	62	118
Bio_resident	70	90	60	100	44	116
Bio_other1	82	93	75	100	58	117
Bio_other2	70	90	60	100	44	116
Bio_other3	76	89	65	100	49	116

**Table 2: Attribute levels for designs with three levels**

Attribute	narrow			base			wide		
Share of forest (%)	-20	0	+20	-25	0	+25	-30	0	+30
Land conversion (%)	-40	0	+40	-50	0	+50	-60	0	+60
Bio_whole	80	85	90	70	85	100	53	85	117
Bio_agrar	76	82	89	65	80	100	49	82	116
Bio_forest	85	90	95	80	90	100	62	90	118
Bio_resident	70	80	90	60	80	100	44	80	116
Bio_other1	82	87	93	75	85	100	58	87	117
Bio_other2	70	80	90	60	80	100	44	80	116
Bio_other3	76	82	89	65	80	100	49	82	116

**Table 3: Attribute levels for designs with 4 levels**

Attribute	narrow				base				wide			
Share of forest (%)	-20	-8	+8	+20	-25	-10	+10	+25	-30	-12	+12	+30
Land conversion (%)	-40	-20	+20	+40	-50	-25	+25	+50	-60	-30	+30	+60
Bio_whole	80	83	87	90	70	80	90	100	53	74	95	117
Bio_agrar	76	80	84	89	65	77	88	100	49	71	93	116
Bio_forest	85	88	91	95	80	86	94	100	62	80	98	118
Bio_resident	70	76	83	90	60	73	87	100	44	68	92	116
Bio_other1	82	85	89	93	75	83	91	100	58	77	96	117
Bio_other2	70	76	83	90	60	73	87	100	44	68	92	116
Bio_other3	76	80	84	89	65	77	88	100	49	71	93	116

## Appendix 2

**Table 1: Linear regression number of drop-outs and design dimensions**

<b>Dimension</b>	<b>Coefficient</b>	<b> t-value </b>
Number of choice sets	0.6344	2.52
Number of alternatives	1.2264	0.65
Number of attributes	4.2485	2.59
Number of levels	-2.2349	1.09
Wide level range	5.0965	1.26
Narrow level range	2.1755	0.57
Constant	-14.9163	-0.99

Note: n = 16

**Table 2: Heteroskedastic logit with order effects CS<sub>2</sub>-CS<sub>24</sub>**

Variable	Coefficient	t-Val.	Variable	Coefficient	t-Val.
ASCsq	0.1842	3.08	CS <sub>1</sub>	0.0000	fixed
Share of forest	0.0065	3.11	CS <sub>2</sub>	-0.0243	0.32
Land conversion	-0.0034	3.12	CS <sub>3</sub>	0.0662	0.90
Bio_whole	0.0043	3.03	CS <sub>4</sub>	0.0561	0.76
Bio_agrar	0.0021	2.89	CS <sub>5</sub>	0.1601	2.24
Bio_forest	0.0024	2.77	CS <sub>6</sub>	0.1818	2.57
Bio_urban	0.0026	2.69	CS <sub>7</sub>	0.1518	1.93
Bio_other1	0.0027	2.83	CS <sub>8</sub>	0.1020	1.26
Bio_other2	0.0010	2.22	CS <sub>9</sub>	0.2437	3.18
Bio_other3	0.0019	2.58	CS <sub>10</sub>	0.2700	3.56
Cost	-0.0023	3.15	CS <sub>11</sub>	0.1905	2.45
Age group 31 to 40 years	-0.0134	0.34	CS <sub>12</sub>	0.1967	2.53
Age group 41 to 50 years	-0.0781	2.08	CS <sub>13</sub>	0.1677	1.89
Age group 51 to 60 yeras	0.0629	1.61	CS <sub>14</sub>	0.2737	3.17
Age group older than 60 years	-0.1647	3.48	CS <sub>15</sub>	0.1618	1.82
Gender male	-0.1287	4.90	CS <sub>16</sub>	0.2460	2.85
Middle-school	0.0049	0.08	CS <sub>17</sub>	0.1549	1.73
High-school	0.1015	1.71	CS <sub>18</sub>	0.2334	2.72
University degree	0.1489	2.63	CS <sub>19</sub>	0.2367	2.13
Dummy 4 alternatives	0.0910	2.21	CS <sub>20</sub>	0.2640	2.41
Dummy 5 alternatives	0.0401	1.14	CS <sub>21</sub>	0.3799	3.63
Number of attributes	0.1496	2.66	CS <sub>22</sub>	0.3802	3.70
Number of levels	0.3861	4.04	CS <sub>23</sub>	0.2846	2.59
Interaction attributes and levels	-0.0784	4.24	CS <sub>24</sub>	0.1851	1.60
Dummy narrow range	0.2318	6.51			
Dummy wide range	-0.0995	2.65			
Log-likelihood null	-31114.18				
Log-likelihood model	-27375.53				
Observations	90354				

## **Danksagung**

An dieser Stelle möchte ich meinen Betreuern und Förderern danken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Professor Dieter danke ich für die Unterstützung während der Arbeit an dieser Dissertation und dafür, dass er zum passenden Zeitpunkt die passenden kritischen Fragen gestellt hat.

Professor Möhring möchte ich herzlich für seine Hilfe und die Übernahme des Zweitgutachtens danken.

Professor Olschewski danke ich für die spontane Bereitschaft, mündlicher Prüfer zu sein. Ausdrücklich danken möchte ich auch all meinen Koautoren für die tolle Zusammenarbeit und die Diskussionen an so vielen Orten, an denen wir gemeinsam waren: Dr. Peter Elsasser, Dr. Jürgen Meyerhoff, Malte Oehlmann und Professor Petr Mariel.

Darüberhinaus danke ich allen Kolleginnen und Kollegen vom Thünen-Institut für die herzliche Arbeitsatmosphäre und die Hilfsbereitschaft, ganz besonders den Kolleginnen aus der Mittagspause.

Ein besonders großer Dank gilt meiner Familie und meinem Mann Till, die immer an mich geglaubt und mich besonders im letzten Jahr der Arbeit unendlich unterstützt haben.

## **Eidesstattliche Erklärung**

Hiermit erkläre ich an Eides statt, dass ich die beigefügte Dissertation selbständig und ohne unerlaubte Hilfe angefertigt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe.

Ich versichere außerdem, dass ich mich vor diesem Promotionsverfahren keiner weiteren Promotionsprüfung unterzogen oder mich zu einer solchen Prüfung angemeldet habe.

Hamburg, den 05.06.2018

---

Priska Weller