



# DIE MÖGLICHEN BESTIMMUNGSFAKTOREN DES ÖLKÜRBISANBAUS IN ÖSTERREICH – EINE RÄUMLICH-ÖKONOMETRISCHE ANALYSE

## Masterarbeit

zur Erlangung des akademischen Grades Diplom-Ingenieur  
im Rahmen des Studiums Agrar- und Ernährungswirtschaft

Eingereicht von:

**Andreas Niedermayr**

Matrikelnummer: 0740688

Betreuer:

**Univ.Prof. Dr. Jochen Kantelhardt**

**Univ. Ass. Dr. Martin Kapfer**

Institut für Agrar- und Forstökonomie

Department für Wirtschafts- und Sozialwissenschaften

Wien, 2014



# Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei der Verfassung meiner Masterarbeit und während meiner Studienzeit unterstützt haben.

Mein besonderer Dank gilt Herrn Univ. Prof. Dr. Jochen Kantelhardt und Herrn Univ. Ass. Dr. Martin Kapfer für die kompetente Betreuung meiner Masterarbeit, die angenehme und produktive Atmosphäre bei Besprechungen sowie ihre wertvollen Anregungen und Ratschläge. Ohne ihre Unterstützung wäre diese Masterarbeit nicht zustande gekommen.

Herzlich bedanken möchte ich mich auch bei meinen Kollegen Stefan Kirchwegger, Tobias Moser, Martin Franzel und meiner Kollegin Lena Luise Schaller. Sie waren mir im Verlauf meiner Arbeit stets bei Fragen behilflich und zu angeregten Diskussionen bereit.

Für die ausführlichen fachlichen Ratschläge und Informationen zum Ölkürbis möchte ich mich bei Herrn DI Dr. Anton Brandstetter (LK Niederösterreich), Herrn Ing. Mag. Andreas Cretnik (Geschäftsführer Gemeinschaft Steirisches Kürbiskernöl g.g.A.) sowie Herrn Ao. Univ. Prof. Dipl.-Ing. Dr. nat. techn. Peter Liebhard bedanken.

Herr Dr. rer. nat. Andreas Meyer-Aurich, Herr Prof. Dipl.-Ing. Dr. techn. Gregor Laaha und Herr Ass. Prof. Dr. Herbert Formayer unterstützten mich bei Fragestellungen zur Methodik. Auch ihnen möchte ich an dieser Stelle meinen Dank aussprechen.

Abschließend möchte ich mich noch bei meiner Familie, allen voran meinen Eltern, bedanken. Ohne ihre fortwährende Unterstützung wäre es mir nicht möglich gewesen, mein Studium erfolgreich abzuschließen.

Bei meinen Freunden und ganz besonders meiner Freundin möchte ich mich noch für den Rückhalt sowie die vielen aufbauenden Worte und Ermunterungen, speziell gegen Ende meines Studiums, bedanken.

Andreas Niedermayr

## **Eidesstattliche Erklärung**

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Masterarbeit selbständig angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

---

Unterschrift

## Kurzfassung

Der Ölkürbisanbau in Österreich hat sich in den letzten Jahren dynamisch entwickelt. Wurde im Jahr 2000 noch auf 10.376 ha Ölkürbis angebaut, so stieg die Anbaufläche bis zum Jahr 2011 um mehr als 100 % auf insgesamt 26.119 ha an.

Im Rahmen der vorliegenden Masterarbeit wird diese dynamische Entwicklung mit unterschiedlichen Regressionsmodellen untersucht, um mögliche Bestimmungsfaktoren des Ölkürbisanbaus in Österreich zu identifizieren. Besonderes Augenmerk wird dabei auf die Rolle räumlicher Agglomerationseffekte gelegt, die im Ölkürbisanbau durch Erfahrungs- und Wissensaustausch, überbetriebliche Maschinenkooperationen und räumliche Bindung zu be- und verarbeitenden Betrieben auftreten können. Räumliche Effekte, wie z. B. Agglomerationseffekte, können mit räumlichen Regressionsmodellen modelliert werden. Daher werden neben einem OLS- und einem Tobit-Modell auch räumliche Regressionsmodelle in Form eines SAR- sowie eines SAR-Tobit Modells berechnet. Die empirische Untersuchung erstreckt sich über 1.347 österreichische Gemeinden, die mit Querschnittsdaten für das Jahr 2011 analysiert werden.

Die Ergebnisse zeigen, dass das SAR-Tobit-Modell die Daten am besten beschreibt, da es sowohl dem hohen Anteil an Nullwerten in der Zielvariable Rechnung trägt, als auch vorhandene räumliche Autokorrelation berücksichtigt. Darüber hinaus zeigt sich, dass der Ölkürbisanbau neben klimatischen Bedingungen sehr stark durch die geographisch geschützte Angabe (g.g.A.) von steirischem Kürbiskernöl und vermutlich auch durch räumliche Agglomerationseffekte beeinflusst wird.

Die Berücksichtigung räumlicher Effekte ist für korrekte Modellergebnisse von zentraler Bedeutung. Darüber hinaus erweitern räumliche Regressionsmodelle die Möglichkeiten der Ergebnisinterpretation, was zu einer vielschichtigen und somit realitätsnäheren Problemanalyse beiträgt. Daher soll mit dieser Masterarbeit auch zur stärkeren Auseinandersetzung mit räumlichen Tobit-Modellen in der Agrarökonomie angeregt werden.

## **Abstract**

The oil pumpkin cultivation in Austria has undergone a dynamic development in recent years. The area under cultivation increased from 10.376 ha in 2000 by over 100% to 26.119 ha in 2011.

In this master thesis different regression models are used to investigate this dynamic development with the goal of identifying possible determinants of the oil pumpkin cultivation in Austria. Particular attention is paid to the role of spatial agglomeration effects, which may occur in oil pumpkin cultivation due to exchange of knowledge and experience, shared use of machinery or the need of spatial proximity to manufacturing and processing companies. Spatial effects, e.g. in form of agglomeration effects, may be modelled with spatial econometric models. Therefore, in addition to an OLS model and a Tobit model, the fitted models also include spatial econometric models in form of a SAR and SAR Tobit specification. The empirical analysis uses a cross sectional data set of 1.347 Austrian communities for the year 2011.

The results indicate that the SAR Tobit model fits the data best, as it takes into account the high proportion of zeroes in the dependent variable as well as existing spatial autocorrelation. Moreover, it can be shown that the oil pumpkin cultivation is strongly influenced by climatic conditions, the protected geographical indication (PGI) of Styrian pumpkin seed oil and probably also by spatial agglomeration effects.

The incorporation of spatial effects is of central importance for correct model results. Furthermore spatial econometric models expand the possibilities of model interpretation, contributing to a more complex and more realistic problem analysis. Therefore, this master thesis also aims to stimulate a more vital discussion on the use of spatial Tobit models in the field of agricultural economics.

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b> .....  | <b>1</b>  |
| <b>2</b> | <b>Der Ölkürbisanbau in Österreich</b> .....                         | <b>4</b>  |
| 2.1      | Standortansprüche, Fruchtfolge und Konkurrenzkulturen .....          | 4         |
| 2.2      | Anbau, Ernte und Verarbeitung.....                                   | 5         |
| 2.3      | Herkunftsschutz von Kürbiskernöl.....                                | 6         |
| 2.4      | Nutzungsarten von Kürbiskernen, Produktion und Vermarktung .....     | 6         |
| 2.5      | Wirtschaftlichkeit des Ölkürbisanbaus.....                           | 7         |
| 2.6      | Zeitliche Entwicklung und räumliche Verteilung des Anbaus.....       | 9         |
| <b>3</b> | <b>Material und Methoden</b> .....                                   | <b>14</b> |
| 3.1      | Untersuchungsbereich.....  | 14        |
| 3.2      | Modellvariablen .....  | 16        |
| 3.2.1    | Abhängige Variable .....   | 16        |
| 3.2.2    | Unabhängige Variablen .....  | 17        |
| 3.2.3    | Deskriptive Statistik der Modellvariablen.....                       | 22        |
| 3.3      | Ökonometrische Modelle und statistische Tests .....                  | 23        |
| 3.3.1    | Lineares Regressionsmodell.....                                      | 24        |
| 3.3.2    | Tobit-Modell.....  | 26        |
| 3.3.3    | Räumliche Regressionsmodelle.....                                    | 28        |
| 3.3.4    | Modellgleichung und Interpretation der Regressionskoeffizienten..... | 33        |
| <b>4</b> | <b>Ergebnisse</b> .....  | <b>38</b> |
| 4.1      | Regressionskoeffizienten, Gütemaße und statistische Tests .....      | 38        |
| 4.2      | Marginale Effekte.....   | 44        |
| 4.3      | Vergleich der gefitteten Werte und Residuen.....                     | 49        |
| <b>5</b> | <b>Diskussion und Schlussfolgerungen</b> .....                       | <b>55</b> |
| 5.1      | Methode .....  | 55        |
| 5.2      | Ergebnisse .....   | 57        |
| 5.3      | Weiterer Forschungsbedarf .....                                      | 62        |

|  |           |
|--|-----------|
| <b>Literatur</b> .....   | <b>63</b> |
| <b>Anhang</b> .....  | <b>68</b> |
| A.    Detaillierte Ergebnistabellen der Regressionsmodelle ..... | 68        |
| B.    R-Skript zur Berechnung der Regressionsmodelle.....        | 75        |



## Abbildungsverzeichnis

|  |    |
|--|----|
| Abbildung 1: Vergleich der Spezialkostenfreien Leistung (DB – kulturspezifische Fixkosten) in €/ha verschiedener Ackerkulturen für die Jahre 2010 und 2011 (basierend auf Arbeitskreisergebnissen im niederösterreichischen Trockengebiet) ..... | 8  |
| Abbildung 2: Entwicklung von Ertrag, Anbaufläche, Gesamtproduktion und Verkaufspreis des Ölkürbis in Österreich von 2000 bis 2011 .....  | 10 |
| Abbildung 3: Entwicklung der Ölkürbis-anbauflächen in Österreich nach Bundesländern von 2000 bis 2011 in ha .....  | 11 |
| Abbildung 4: Räumliche Verteilung des Ölkürbis-anbaus in Oberösterreich, Niederösterreich, Burgenland, Steiermark und Kärnten 2000 (N = 1347 Gemeinden).....   | 13 |
| Abbildung 5: Räumliche Verteilung des Ölkürbis-anbaus in Oberösterreich, Niederösterreich, Burgenland, Steiermark und Kärnten 2011 (N = 1347 Gemeinden).....   | 13 |
| Abbildung 6: Lage der Gemeinden des Untersuchungsgebietes .....  | 15 |
| Abbildung 7: Räumliche Darstellung der Dummy Variablen.....  | 18 |
| Abbildung 8: Flow-Chart der Aufbereitung der Temperaturdaten.....  | 20 |
| Abbildung 9: Diagnostikplots OLS-Modell.....   | 43 |
| Abbildung 10: Diagnostikplots SAR-Tobit-Modell.....  | 44 |
| Abbildung 11: Ölkürbisanteile in Prozent der Ackerfläche nach Gemeinden, Jahr 2011 .....   | 49 |
| Abbildung 12: Geschätzte Ölkürbisanteile in Prozent der Ackerfläche nach Gemeinden, Jahr 2011 .....  | 52 |
| Abbildung 13: Abweichungen der Schätzer von den Ölkürbisanteilen in Prozent der Ackerfläche nach Gemeinden, Jahr 2011 .....  | 54 |

# Tabellenverzeichnis

|   |    |
|---|----|
| Tabelle 1: Zusammenfassung der Modellvariablen – Hypothesen und deskriptive Statistik<br>(n=1347 Gemeinden) .....   | 23 |
| Tabelle 2: Ergebnisse der Regressionsmodelle: Regressionskoeffizienten, Gütemaße und<br>Teststatistiken .....   | 39 |
| Tabelle 3: Marginale Effekte Tobit-Modell.....  | 46 |
| Tabelle 4: Marginale Effekte SAR-Tobit-Modell (ADI, All und ATI) .....  | 48 |
| Tabelle 5: Geschätzter Flächenbedarf für den Ölkürbisanbau 2011 .....   | 53 |
| Tabelle 6: Ergebnisse OLS-Modell .....  | 68 |
| Tabelle 7: Marginale Effekte OLS-Modell .....   | 69 |
| Tabelle 8: Ergebnisse Tobit-Modell .....  | 70 |
| Tabelle 9: Zusammenfassung unterschiedlicher Nachbarschaftsmatrizen, basierend auf<br>kontinuierlicher Nachbarschaft bzw. Schwellenentfernungen mit inverser<br>Distanzgewichtung ..... | 71 |
| Tabelle 10: Ergebnisse SAR-Modell .....   | 72 |
| Tabelle 11: Marginale Effekte SAR-Modell .....  | 73 |
| Tabelle 12: Ergebnisse SAR-Tobit-Modell .....   | 74 |

## Abkürzungsverzeichnis

|        |  |
|--------|--|
| ADI    | Average Direct Impact  |
| AIC    | Akaike information criterion   |
| All    | Average Indirect Impact  |
| Akh    | Arbeitskraftstunden  |
| AMA    | Agrarmarkt Austria   |
| APE    | Average Partial Effect   |
| ATI    | Average Total Impact   |
| BLUE   | Best Linear Unbiased Estimator   |
| BMLFUW | Bundesministerium für Landwirtschaft, Forstwirtschaft, Umwelt und Wasserwirtschaft |
| ca.    | cirka  |
| CORINE | Coordination of Information on the Environment                                     |
| g.g.A. | geographisch geschützte Angabe   |
| GIS    | geographisches Informationssystem  |
| GVE    | Großvieheinheit  |
| LF     | Landwirtschaftlich genutzte Fläche   |
| ML     | Maximum Likelihood   |
| OLS    | Ordinary Least Squares   |
| SAR    | Spatial Autoregressive Model   |
| SARAR  | Spatial Autoregressive Model with Autoregressive Disturbances                      |
| SARMA  | Spatial Autoregressive Moving Average Model  |
| SEM    | Spatial Error Model  |
| Tobit  | Tobin's Probit Model   |
| usw.   | und so weiter  |
| W      | Nachbarschaftsmatrix   |
| z. B.  | zum Beispiel   |

# 1 Einleitung

Die räumliche Verteilung von Kulturpflanzen im Ackerbau wird von vielen Faktoren beeinflusst. Hierzu zählen etwa natürliche Standortfaktoren (z.B. Niederschlag, Temperatur, Boden oder Topographie), betriebliche Standortfaktoren (z.B. verfügbare Ackerfläche, Maschinenausstattung, Nähe zu be- und verarbeitenden Betrieben, bisherige betriebliche Ausrichtung) oder ackerkulturabhängige Faktoren (notwendige spezielle Mechanisierung, Arbeitsaufwand, Deckungsbeitrag, Anforderungen an die Fruchtfolge). Der Entscheidungsprozess der Kulturauswahl im Ackerbau gestaltet sich durch die vielen Einflussfaktoren und die zahlreichen möglichen Wechselwirkungen zwischen den Faktoren äußerst komplex. Laut ODGAARD et al. (2011, 291) sind die exakten Mechanismen, welche die räumliche Verteilung von Ackerkulturen beeinflussen, im Detail bisher noch wenig erforscht.

Diese Arbeit befasst sich mit dem genannten Themenbereich. Es werden mögliche Bestimmungsfaktoren der räumlichen Verteilung des Ölkürbisanbaus in Österreich untersucht. Der Ölkürbisanbau hat sich in den letzten Jahren dynamisch entwickelt. Wurde im Jahr 2000 noch auf 10.376 ha Ölkürbis angebaut, so stieg die Anbaufläche bis zum Jahr 2011 um mehr als 100 % auf insgesamt 26.119 ha an (BMLFUW 2012, s.p.). Der Anbau nahm vor allem sowohl im traditionellen Anbaugebiet Steiermark als auch in Niederösterreich und dem Burgenland überproportional zu. In diesen Bundesländern herrschen nicht nur günstige klimatische Bedingungen für den Ölkürbisanbau, sondern es liegt dort auch das Herkunftsgebiet für „Steirisches Kürbiskernöl g.g.A.“ (BMLFUW s.a., 3).

Ziel dieser Arbeit ist es, diese Entwicklung näher zu betrachten, um mögliche Bestimmungsfaktoren und daraus abgeleitet weitere Potentiale des Ölkürbisanbaus in Österreich zu identifizieren. Im Vordergrund der Analyse steht dabei die Untersuchung räumlicher Effekte. Aus diesen Überlegungen ergeben sich 3 Forschungsfragen, die im Verlauf dieser Arbeit beantwortet werden sollen. Die erste Frage lautet *„Welche Bestimmungsfaktoren haben einen statistisch signifikanten Einfluss auf den Ölkürbisanbau in ausgewählten Gebieten Österreichs?“*. Die beiden anderen Fragen bauen auf dieser zentralen Fragestellung auf. Mit der Frage *„Wie wirken die statistisch signifikanten Bestimmungsfaktoren auf den Ölkürbisanbau (positiv oder negativ)?“* wird näher auf die Wirkungsweise der Bestimmungsfaktoren eingegangen. Die Beantwortung der letzten Frage soll schließlich die Rolle möglicher räumlicher Effekte klären. Sie lautet *„Haben räumliche Effekte einen statistisch nachweisbaren Einfluss auf den Ölkürbisanbau?“*.

Als methodischer Zugang in der vorliegenden Arbeit wird die räumliche Regressionsanalyse gewählt. Dabei werden herkömmliche Regressionsmodelle in einer Weise erweitert, dass

mögliche räumliche Abhängigkeit (räumliche Autokorrelation) in den Modellberechnungen erfasst werden kann. Räumliche Autokorrelation bedeutet, dass sich die Werte räumlich naher Beobachtungen in ihren Ausprägungen ähneln (PLANT 2012, 59). Das bedeutet, was an einem Ort passiert, beeinflusst Entwicklungen an einem benachbarten Ort. Wird vorhandene räumliche Autokorrelation in einem Regressionsmodell nicht berücksichtigt, so kann dies zu verzerrten Schätzungen der Regressionsparameter führen (LANGYINTUO und MEKURIA 2008, 153).

Die räumliche Regressionsanalyse wird im landwirtschaftlichen Umfeld auf wissenschaftlicher Ebene vielfach angewendet, etwa in den Bereichen „Precision Farming“ (siehe z.B. ANSELIN et al. 2004 oder MEYER-AURICH et al. 2010), „regionale Konvergenz“ (siehe z.B. BIVAND und BRUNSTAD 2006 oder FINGLETON und LÓPEZ-BAZO 2006) oder „Landnutzungsmodellierung“ (siehe z.B. AGUIAR et al. 2007, GELLRICH und ZIMMERMANN 2006 oder BRADY und IRWIN 2011). Die Rolle von räumlicher Autokorrelation bei der Erklärung der räumlichen Verteilung von Kulturpflanzen im Ackerbau wird jedoch bisher nur in einzelnen Studien berücksichtigt. GRIFFITH (2008) befasst sich mit dem Zuckerrohranbau in Puerto Rico, ODGAARD et al. (2011) sowie LIPPERT et al. (2012) untersuchen Faktoren, welche die Maisanbaufläche in Dänemark bzw. Deutschland (Baden-Württemberg) beeinflussen und SCHOLZ et al. (2013) beschäftigen sich mit den Bestimmungsfaktoren des Silomaisanbaus in Bayern.

Die räumliche Regressionsanalyse wird außerdem beinahe ausschließlich in Fällen angewandt, in denen die abhängige Variable des Regressionsmodells kontinuierlich ist. In der Praxis treten aber häufig Situationen auf, in denen ein Grenzwert die Ausprägungen der abhängigen Variable einschränkt. So wird etwa in vielen österreichischen Gemeinden Ölkürbis in unterschiedlichem Umfang angebaut, in zahlreichen Gemeinden wird aber kein Ölkürbis kultiviert. In solchen Fällen empfiehlt sich die Anwendung eines Tobit-Modells (WOOLDRIDGE 2009, 587f), welches durch die Berücksichtigung etwaiger räumlicher Autokorrelation zu einem räumlichen Tobit-Modell erweitert werden kann (LESAGE und PACE 2009, 299). Im agrarökonomischen Bereich verwenden z. B. LANGYINTUO und MEKURIA (2008) ein räumliches Tobit-Modell, um räumliche Abhängigkeit bei der Adaption neuer Maissorten durch Landwirte und Landwirtinnen in Mosambik zu untersuchen.

In anderen Wissenschaftszweigen finden räumliche Tobit-Modelle ebenfalls vereinzelt Verwendung. So untersuchen AUTANT-BERNARD und LESAGE (2009) etwa räumliche Abhängigkeit im Wissensbildungsprozess unterschiedlicher französischer Industriesektoren, der die Generierung von Patenten zur Folge hat. DONFOUET et al. (2013) beschäftigen sich mit der Rolle von räumlicher Autokorrelation bei der Nachfrage nach einer kommunalen Gesundheitsversicherung in Entwicklungsländern.

Mit der vorliegenden Arbeit soll ein Beitrag zum wissenschaftlichen Diskurs in diesem Bereich geleistet werden. Anhand des Fallbeispiels „Ölkürbisanbau in Österreich“, werden neben einem herkömmlichen linearen Regressionsmodell und dessen räumlicher Erweiterung, außerdem ein Tobit Modell sowie ein räumliches Tobit Modell gebildet.

Der Aufbau der Arbeit gliedert sich folgendermaßen: Im 2. Kapitel erfolgt zunächst eine theoriebasierte Beschreibung des Ölkürbisanbaus in Österreich. Neben den natürlichen Standortansprüchen des Ölkürbis sowie der Produktions- und Vermarktungsstruktur, wird näher auf die räumliche und zeitliche Entwicklung des Ölkürbisanbaus eingegangen. In Kapitel 3 werden zunächst der räumliche und zeitliche Untersuchungsbereich der Arbeit festgelegt. Danach werden, aufbauend auf den Erkenntnissen aus Kapitel 2, die Variablen für die Regressionsmodelle ausgewählt. Schließlich werden die verwendeten Regressionsmodelle und statistischen Testverfahren erläutert. In Kapitel 4 sind die Ergebnisse der Arbeit dargestellt, wobei die unterschiedlichen Modelle miteinander verglichen werden. In Kapitel 5 werden die Ergebnisse und die angewandte Methodik, unter Bezugnahme auf relevante Literatur kritisch reflektiert und weiterer Forschungsbedarf aufgezeigt.

## 2 Der Ölkürbisanbau in Österreich

Weltweit werden jährlich auf rund 600.000ha ungefähr 200.000 t Ölkürbiskerne geerntet. Neben Asien (120.000 t) konzentriert sich der Ölkürbisanbau auch auf Europa (80.000 t) (BUCHER 2010, s.p., zit. nach DÖRFLER 2011, 18). Rund ein Fünftel der europäischen Ölkürbisernte stammt aus Österreich. Neben den günstigen klimatischen Bedingungen hat der Ölkürbisanbau in Österreich, speziell in der Steiermark eine lange Tradition.

In diesem Kapitel wird der Ölkürbisanbau in Österreich beschrieben. Zu Beginn werden seine natürlichen Standortansprüche erläutert und näher auf den Anbau, die Ernte, Verarbeitung und Wirtschaftlichkeit eingegangen. Im Anschluss daran wird die zeitliche Entwicklung des Ölkürbisanbaus in Österreich näher beleuchtet. Dabei wird besonderes Augenmerk auf die räumliche Verteilung der Anbauflächen und deren Entwicklung gelegt. Die Erkenntnisse aus diesem Kapitel dienen als Basis für die Bildung möglicher Bestimmungsfaktoren des Ölkürbisanbaus in Österreich.

### 2.1 Standortansprüche, Fruchtfolge und Konkurrenzkulturen

Der Ölkürbis ist sehr wärmebedürftig. Für die Auskeimung und das Wachstum benötigt er eine Mindesttemperatur von rund 10°C (HEYLAND et al. 2006, 167). Die Jahresdurchschnittstemperatur für den Ölkürbisanbau sollte mindestens 8°C betragen (KULMER und WEBER 1990, 14). Aufgrund fehlender Frostresistenz benötigt der Ölkürbis außerdem während seines Wachstums (Saat Ende April – Anfang Mai, Ernte meist bis Ende September) Frostfreiheit (KULMER und WEBER 1990, 14). Bei der Einschätzung der Temperaturansprüche des Ölkürbis orientiert man sich in der Praxis am Körnermais (BRANDSTETTER 2013, s.p.).

Für gute Kernerträge sind vor allem ausreichende Niederschläge zwischen Ende Juni und Mitte August notwendig (KULMER und WEBER 1990, 14). Daher ist der Ölkürbisanbau speziell im warm-feuchten Illyrischen Klimagebiet der südöstlichen Steiermark und dem südlichen Burgenland weit verbreitet (SPERBER 1988, 48). Nasskalte Witterung und Wind wirken hingegen ertragsmindernd (KULMER und WEBER 1990, 14f). So führen etwa zu hohe Niederschläge während der Vegetationsperiode zu einem hohen Fäulnisanteil (CRETNIK 2013, s.p.).

An den Boden stellt der Ölkürbis keine hohen Ansprüche (DIEPENBROCK 1999, 314), wobei er auf leichten bis mittelschweren, humusreichen Böden mit guter Wasserführung und Struktur (sandige Lehm- bis lehmige Sandböden) am besten gedeiht (KULMER und WEBER 1990, 15). Solche Böden bieten aufgrund ihrer gröberen Körnung außerdem den Vorteil, dass sie sich

rascher erwärmen (DACHLER und PELZMANN 1999, s.p. zit. nach HEYLAND et al. 2006, 167). Da der Ölkürbis auf Bodenverdichtung empfindlich reagiert, sind schlecht strukturierte, schwere, tonige Böden nicht für den Anbau geeignet (SCHUSTER 1977, 19).

An die Vorfrucht stellt der Ölkürbis grundsätzlich keine besonderen Ansprüche, sodass alle Feldfrüchte geeignet sind, sofern sie keine Fungizid-, Herbizid- oder Insektizidrückstände im Boden hinterlassen (SPERBER 1988, 48). Als Vorfrüchte eignen sich besonders Gras- und Klee grasumbrüche sowie alle gesunden Getreidefelder (NEUBAUER 2009, 15). Ungünstige Vorfrüchte aufgrund belasteter Ernterückstände sind Raps, Feldgemüse, Sonnenblume, Ackerbohne, Zuckerrübe oder Kartoffeln (NEUBAUER 2009, 15f). Gurke, Zucchini und Kürbis scheiden ebenfalls als Vorfrüchte aus, da sie das Zucchini gelbmosaikvirus übertragen können (LIEBHARD et al, 1999, 24, zit. nach SCHWARZ 2008, 17). HEYLAND (2006, 167) rät aus phytopathologischen Gründen zu einer Anbaupause von 4 Jahren zwischen zwei Ölkürbiskulturen. Konkurrenzulturen zum Ölkürbis sind prinzipiell alle Sommerungen (BRANDSTETTER 2013, s.p.). In der Steiermark sind dabei besonders Körnermais und Sojabohne von Bedeutung (BRANDSTETTER 2013, s.p.).

### **2.2 Anbau, Ernte und Verarbeitung**

Der großflächige Anbau erfolgt wie beim Mais in Reihensaat mittels pneumatischer Einzelkornsämaschine (KULMER und WEBER 1990, 19). Einige Tage vor der Ernte werden die Kürbisse in Reihen zusammengeschoben (SPERBER 1988, 52). Die Ernte erfolgt dann mit speziellen Kürbiserntemaschinen oder umgebauten Mähdreschern (HEYLAND et al. 2006, 170). Die Kerne werden beim Erntevorgang von den Kürbissen getrennt, wobei das Fruchtfleisch meist ungenutzt auf dem Feld verbleibt (HEYLAND et al. 2006, 170). Mit einer Kürbiserntemaschine wird im Schnitt eine Ernteleistung von rund 5 ha pro Tag erreicht (NEUBAUER 2009, 21).

Im Anschluss an die Ernte müssen die Kürbiskerne zuerst gewaschen (meist mit selbstgebauten Waschanlagen) und danach getrocknet werden (in selbst gebauten Trocknungsanlagen oder vollautomatischen Flachrocknungsanlagen) (NEUBAUER 2009, 21). Ernte, Waschung und Trocknung werden vor allem in Niederösterreich überwiegend im Lohnverfahren organisiert (Maschinenring oder private Lohnunternehmer) (BRANDSTETTER 2013, s.p.). Die Investitionen in spezielle Mechanisierung sind für einzelne Betriebe meist nicht rentabel (BRANDSTETTER 2013, s.p.). Da die Kerne in Niederösterreich vereinzelt sogar bis zu 30 km zur nächsten Wasch- und Trocknungsanlage transportiert werden, könnte beim Waschen und Trocknen eine gewisse räumliche Bindung an vorhandene Infrastruktur bestehen (BRANDSTETTER 2013 s.p.).



### **2.3 Herkunftsschutz von Kürbiskernöl**

Für den Begriff „Steirisches Kürbiskernöl“ besteht seit 1998 ein Herkunftsschutz seitens der Europäischen Union (GLAWOGGER 2007, s.p., zit. nach SCHWARZ 2008 42). Für die Umsetzung des Herkunftsschutzes wurde die „Gemeinschaft Steirisches Kürbiskernöl g.g.A.“ (vormals Erzeugerring „Steirisches Kürbiskernöl g.g.A.“) gegründet (SCHWARZ 2008, 1). Neben dieser Aufgabe beschäftigt sich die Gemeinschaft Steirisches Kürbiskernöl g.g.A. mit dem Aufbau der Marke „Steirisches Kürbiskernöl g.g.A.“ und der Vermarktung von „Steirischem Kürbiskernöl g.g.A.“ (CRETNIK 2013, s.p.).

Ölkürbiskerne, die zu „Steirischem Kürbiskernöl g.g.A.“ verarbeitet werden, dürfen nur aus bestimmten Gebieten stammen (BMLFUW s.a., 3). Die definierten Anbaugebiete umfassen die südliche Steiermark (politische Bezirke Deutschlandsberg, Feldbach, Fürstenfeld, Graz-Umgebung, Hartberg, Leibnitz, Radkersburg, Voitsberg und Weiz), das südliche Burgenland (politische Bezirke Jennersdorf, Güssing und Oberwart) sowie Teile von Niederösterreich (politische Bezirke Hollabrunn, Horn, Mistelbach, Melk, Gänserndorf – eingeschränkt auf den Gerichtsbezirk Zistersdorf und Korneuburg-Stockerau – eingeschränkt auf den Gerichtsbezirk Stockerau) (SCHWARZ 2008, 31). Die Verarbeitung der Kerne zu „Steirischem Kürbiskernöl“ darf ausschließlich in den oben genannten Bezirken in der südlichen Steiermark sowie im südlichen Burgenland erfolgen (SPEZIFIKATION STEIRISCHES KÜRBISKERNÖL 1995, zit. nach SCHWARZ 2008, 32). Landwirte und Landwirtinnen bzw. Ölmühlen in den oben genannten Gebieten, können der Gemeinschaft Steirisches Kürbiskernöl g.g.A. beitreten. Sie verpflichten sich mit dem Beitritt einerseits zur Einhaltung spezieller Produktionsauflagen, profitieren aber andererseits auch von den besseren Vermarktungsmöglichkeiten der Marke „Steirisches Kürbiskernöl g.g.A.“ als regionale Spezialität (CRETNIK 2013, s.p.).

### **2.4 Nutzungsarten von Kürbiskernen, Produktion und Vermarktung**

Österreichische Ölkürbiskerne werden zur Gewinnung von Kürbiskernöl (62%), als Knabberkernen (35%) und zur Saatgutproduktion (3%) verwendet (CRETNIK 2013, s.p.). Als Knabberkerne werden Kerne höchster Qualität vermarktet, da auf diesem Weg eine höhere Wertschöpfung erzielt werden kann (CRETNIK 2013, s.p.). Weitere Verwendungsbereiche der Kerne liegen in der Backwarenindustrie und in der Pharmazie. Dafür wird jedoch überwiegend Importware verwendet (CRETNIK 2013, s.p.).

DÖRFLER (2011, 17) bietet, unter Bezugnahme auf die Landwirtschaftskammern Steiermark und Niederösterreich, eine ungefähre Einschätzung des Direktvermarktungsanteils von Kürbiskernöl: Die Direktvermarktung spielt vor allem in der Steiermark eine große Rolle

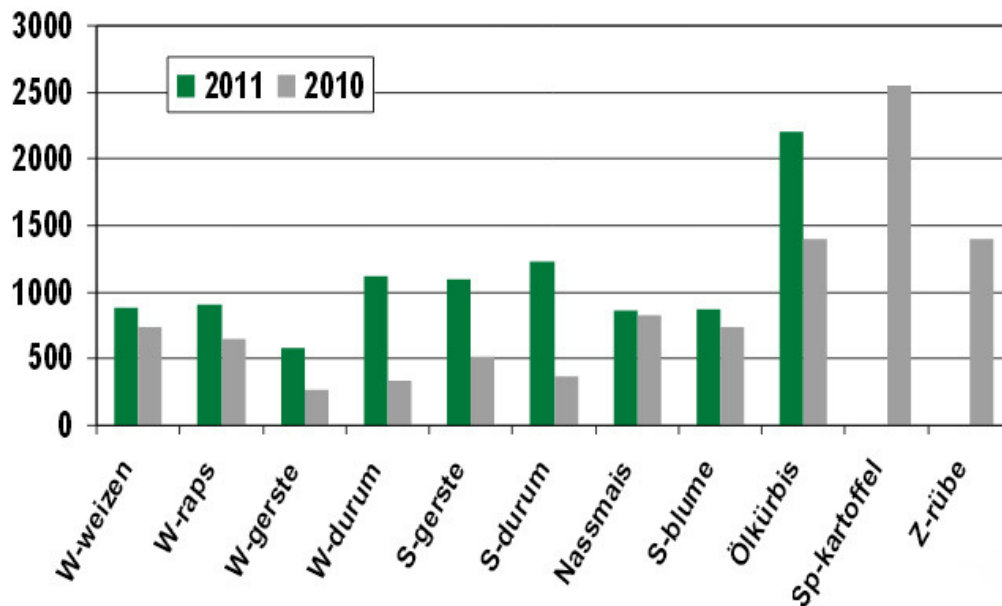
(ca. 60-70%). Die Direktvermarkter pressen das Kernöl selbst oder lassen es teilweise auch im Lohnpressverfahren von nahe gelegenen Ölmühlen pressen, aber den Verkauf übernehmen sie selbst. Überschussware verkaufen sie an Ölmühlen oder andere Direktvermarkter. In Niederösterreich hat die Direktvermarktung im Vergleich dazu einen geringeren Stellenwert (ca. 10-20%). Für niederösterreichische Landwirte bzw. Landwirtinnen spielt der Vertragsanbau mit rund 90% eine große Rolle (BRANDSTETTER 2013, s.p.). Die Kerne für die Kernölproduktion werden im niederösterreichischen g.g.A. Gebiet bei Gemeinschaften gesammelt und dann per LKW-Sammeltransport in die Steiermark bzw. das südliche Burgenland transportiert. Dort erfolgt in den Ölmühlen die Weiterverarbeitung zu Kürbiskernöl (BRANDSTETTER 2013, s.p.). Die Transportkosten spielen dabei aufgrund der Sammeltransporte und des relativ hohen Wertes pro transportierter Gewichtseinheit kaum eine Rolle (BRANDSTETTER 2013, s.p.).

Der überwiegende Anteil der rund 70 österreichischen Ölmühlen, die Kürbiskernöl erzeugen, befindet in der Steiermark sowie teilweise auch im südlichen Burgenland (CRETNIK 2013, s.p.). Die Gemeinschaft Steirisches Kürbiskernöl hat insgesamt 30 Ölmühlen als Mitglieder (CRETNIK 2013, s.p.). Von den großen Ölmühlen sind allerdings nur 3 keine Mitglieder beim Erzeugerring (CRETNIK 2013, s.p.). Der Rest sind sehr kleine, bzw. von Landwirten und Landwirtinnen überwiegend für den Eigenbedarf betriebene Ölmühlen (CRETNIK 2013, s.p.).

### **2.5 Wirtschaftlichkeit des Ölkürbisanbaus**

Aus ökonomischer Perspektive erscheint eine Beurteilung des Ölkürbisanbaus nur im Vergleich mit konkurrierenden Ackerkulturen sinnvoll. In der Anwendung „Internetdeckungsbeitrag“ der Bundesanstalt für Agrarwirtschaft, welche einen Deckungsbeitragsvergleich zwischen verschiedenen Ackerkulturen ermöglicht, finden sich derzeit (Stand 01.03.2014) keine Daten für den Ölkürbis. Nachfolgend sind in Abbildung 1 Daten der Landwirtschaftskammer Niederösterreich dargestellt, die einen Vergleich der spezialkostenfreien Leistungen (Deckungsbeitrag abzüglich kulturspezifischer Fixkosten) des Ölkürbis und anderer Ackerkulturen für die Jahre 2010 und 2011 zeigen. Der Ölkürbis erzielt in beiden Jahren im Vergleich zu den anderen Ackerkulturen eine sehr hohe spezialkostenfreie Leistung. Diese Zahlen sind jedoch nur bedingt aussagekräftig, da sie gebietsbezogen sind (niederösterreichisches Trockengebiet), auf Arbeitskreisergebnissen beruhen (geringe Fallzahl, keine Versuchsergebnisse) und in ihren einzelnen Komponenten (Ertrag, Saatgutkosten, Düngemittelkosten, Pflanzenschutzkosten, variable und kulturspezifische Maschinenkosten) betriebsindividuell stark variieren können. Hinzu kommt, dass eine betriebsinterne Verwendung der Erzeugnisse (z.B. als Futtermittel für die

Veredelung) oder eine etwaige Kuppelproduktnutzung (etwa Strohnutzung bei Getreide) dabei unberücksichtigt bleiben.



**Abbildung 1: Vergleich der Spezialkostenfreien Leistung (DB – kulturspezifische Fixkosten) in €/ha verschiedener Ackerkulturen für die Jahre 2010 und 2011 (basierend auf Arbeitskreisergebnissen im niederösterreichischen Trockengebiet)**

Quelle: Landwirtschaftskammer Niederösterreich (s.a.)

Der Preis für Ölkürbiskerne, hängt laut BRANDSTETTER (2013, s.p.) neben der Nachfragesituation sehr stark davon ab, ob die Ölkürbiskerne nach der Spezifikation „Steirisches Kürbiskernöl g.g.A.“ produziert werden. Da sich „g.g.A. Ölkürbiskerne“ als regionale Spezialität besser vermarkten lassen, liegen deren Preise über denen herkömmlicher Ölkürbiskerne (BRANDSTETTER 2013, s.p.). So betrug der Erzeugerpreis für g.g.A. Ölkürbiskerne 2012 etwa rund 3.000 €/t und für nicht g.g.A. Ölkürbiskerne ca. 2.500 €/t (BRANDSTETTER 2013, s.p.). In einem schwächeren Produktionsjahr sind bei entsprechender Nachfrage die preislichen Unterschiede jedoch geringer oder kaum vorhanden (BRANDSTETTER 2013, s.p.). Die in Abbildung 1 dargestellten Vergleichswerte sind aus den bisher genannten Gründen entsprechend vorsichtig zu interpretieren, zeigen aber durchaus, dass der Ölkürbisanbau aus wirtschaftlicher Sicht eine interessante Alternative für Landwirte und Landwirtinnen darstellen kann.

Der Ölkürbis hat im Vergleich zu etwa Getreide, Mais, Sojabohne oder Raps einen höheren Gesamtarbeitsbedarf (BMLFUW 2008, 84ff): Der Arbeitsaufwand beträgt pro ha Ölkürbisfläche 15-25 Arbeitskraftstunden bei Lohnmaschinen und 20-30 Akh bei eigenen Maschinen. Bei Getreide, Mais, Sojabohne und Raps liegt er im Vergleich dazu im Lohndruschverfahren bei etwa 13-17 AKh/ha und bei eigenen Maschinen bei etwa 17-20

AKh/ha. Der zusätzliche Arbeitsaufwand beim Ölkürbis ergibt sich durch etwa 10 AKh/ha Handhacke (falls notwendig), 1 AKh/ha maschinelles hacken und 0,5 AKh/ha für das Schieben der Kürbisse in Reihen vor der Ernte (BRANDSTETTER 2013, s.p.).

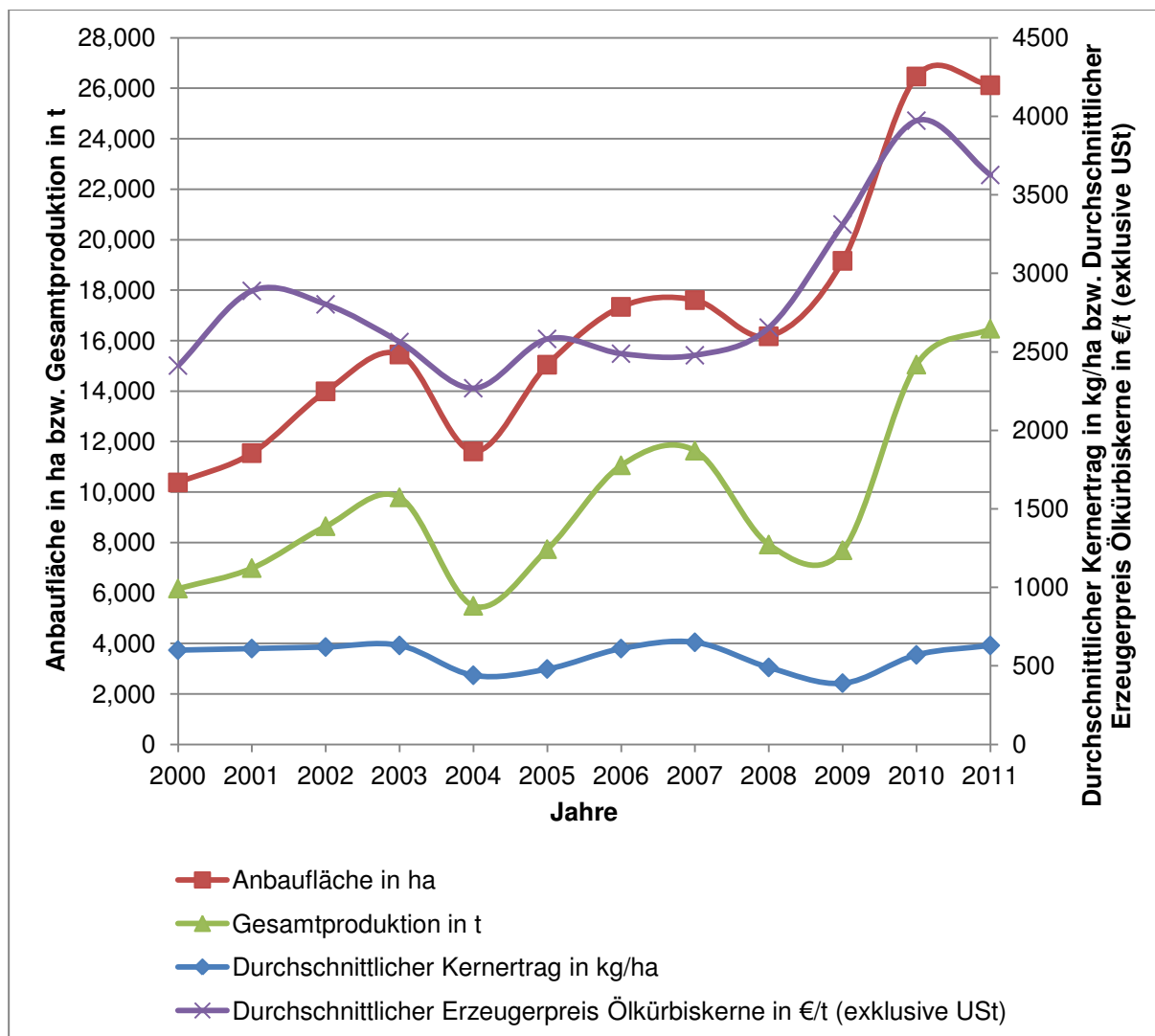
Im Rahmen der ÖPUL-Maßnahme „Integrierte Produktion Gemüse“ kann der Anbau von Ölkürbis mit 250 €/ha gefördert werden (AMA, 2007 s.p.). Voraussetzung dafür ist jedoch unter anderem der Besuch von mindestens 2 Schulungen während des Verpflichtungszeitraums. Eine weitere ÖPUL-Maßnahme, die den Ölkürbisanbau beeinflussen könnte, ist die „Umweltgerechte Bewirtschaftung von Acker- und Grünlandflächen“. Landwirte und Landwirtinnen, die an dieser Maßnahme teilnehmen, dürfen einen maximalen Getreide-Mais-Anteil in der Fruchtfolge von 75% nicht überschreiten (BRANDSTETTER 2013 s.p.). Der Ölkürbis, aber auch beispielsweise Raps, Sojabohne, Sonnenblume, Zuckerrübe etc. stellen für teilnehmende Betriebe eine Alternative dar (BRANDSTETTER 2013 s.p.).

### **2.6 Zeitliche Entwicklung und räumliche Verteilung des Anbaus**

In Abbildung 2 ist die Entwicklung des Ölkürbisanbaus in Österreich von 2000 bis 2011 dargestellt. Neben Ertrag, Anbaufläche und Gesamtproduktion ist auch der durchschnittliche Erzeugerpreis abgebildet. In diesem Zeitraum ist eine positive Entwicklung zu beobachten. Die Anbauflächen und folglich auch die Gesamtproduktion stiegen. Der Preis für Ölkürbiskerne nahm ebenfalls zu.

Der kurzzeitige Rückgang der Anbauflächen im Jahr 2004 beruhte unter anderem auf einer schlechten Preisentwicklung (STATISTIK AUSTRIA 2005, 14). So fiel der Erzeugerpreis für Ölkürbiskerne Anfang 2004 von rund 2.200 €/t im Jänner auf knapp unter 2.000 €/t im Mai (STATISTIK AUSTRIA 2005, 126). Von 2007 bis 2010 nahm der Preis für Ölkürbiskerne kontinuierlich zu. Die Nachfrage nach Steirischem Kürbiskernöl g.g.A. stieg in diesem Zeitraum, da es ab 2006 auch nach Deutschland exportiert wurde (CRETNIK 2013, s.p.). Gleichzeitig führten 2 ertragsschwache Jahre (2008 und 2009) zu einer weiteren Preissteigerung und Flächenausdehnung (CRETNIK 2013, s.p.). 2011 sank der Preis aufgrund einer Rekordernte leicht (CRETNIK 2013, s.p.).

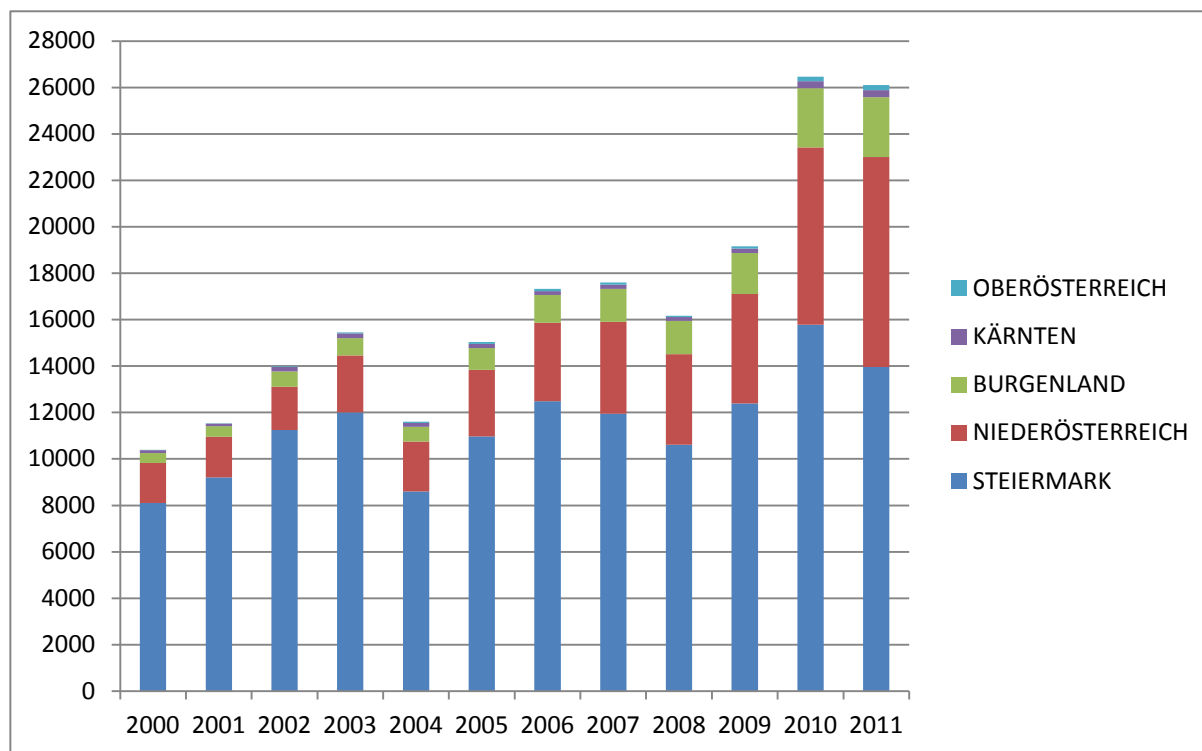
Der Kernertrag variiert in Abhängigkeit von der Anbauregion und den dortigen Witterungsbedingungen stark. Im Durchschnitt konnte österreichweit zwischen 2000 und 2011 ein Ertrag von rund 560 kg/ha erzielt werden.



**Abbildung 2: Entwicklung von Ertrag, Anbaufläche, Gesamtproduktion und Verkaufspreis des Ölkürbis in Österreich von 2000 bis 2011**

Quelle: Eigene Darstellung, Daten aus den Grünen Berichten 2000-2012 (BMLFUW 2001-2012) und der GEDABA (2012)

Abbildung 3 zeigt die Entwicklung der Ölkürbisanbauflächen von 2000 bis 2011 nach Bundesländern. In Vorarlberg, Tirol, Salzburg sowie in Wien befinden sich keine bzw. äußerst geringe Anbauflächen. Deshalb sind diese Bundesländer in der Abbildung nicht angeführt. Die Anbauflächen haben sich im Beobachtungszeitraum mehr als verdoppelt, wobei sie sich speziell in der Steiermark, in Niederösterreich und im Burgenland ausdehnten. Im Zeitraum von 2000 bis 2011 stieg die Ölkürbisanbaufläche in Österreich von rund 10.000 ha auf über 26.000 ha.



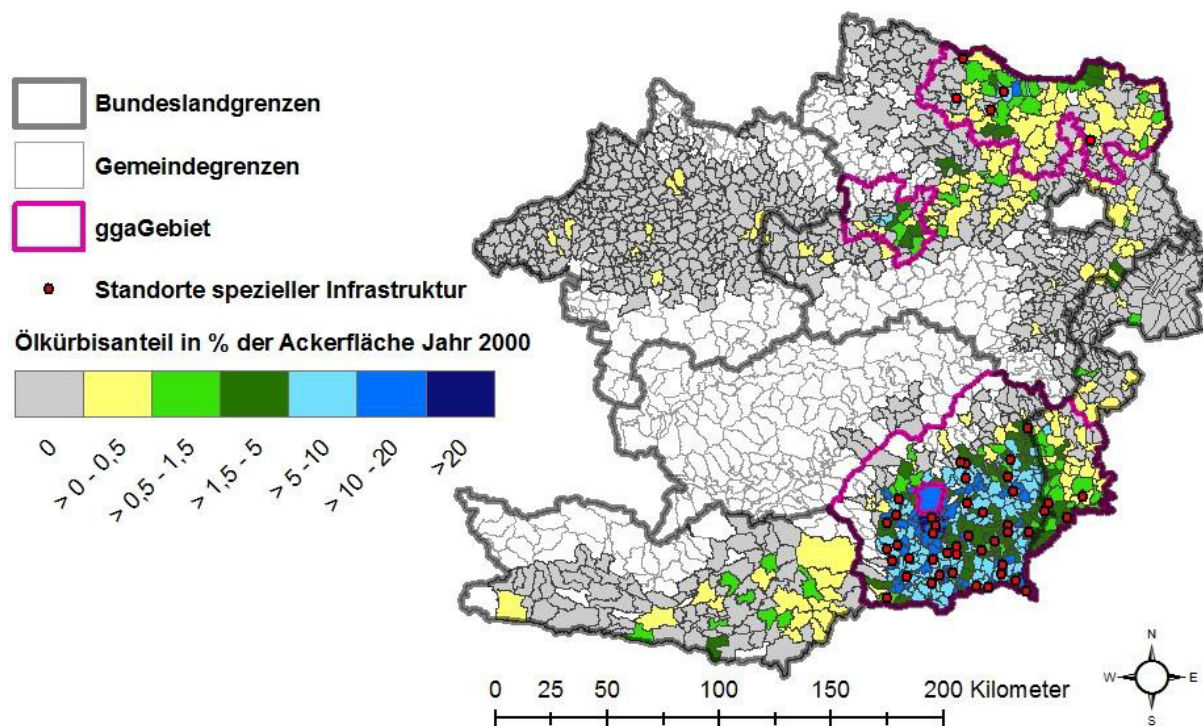
**Abbildung 3: Entwicklung der Ölkürbisanbauflächen in Österreich nach Bundesländern von 2000 bis 2011 in ha**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012)

Eine detailliertere Übersicht der räumlichen Verteilung der Ölkürbisflächen für die Jahre 2000 und 2011 für ausgewählte Gemeinden findet sich in Abbildung 4 und Abbildung 5. Der Ölkürbisanteil in Prozent der Ackerfläche je Gemeinde ist in unterschiedlichen farbigen Kategorien dargestellt. So sind der ersten Kategorie (Grau) z. B. Gemeinden zugeordnet, in denen kein Ölkürbis angebaut wird. Die Gebiete, in denen Ölkürbis nach der Spezifikation „Steirisches Kürbiskernöl g.g.A.“ produziert werden darf (g.g.A. Gebiet), sind durch violette Umrahmung gekennzeichnet. Die als Punkte eingezeichneten Standortdaten von insgesamt 43 Ölmühlen in der Steiermark und dem Burgenland, wurden von der „Gemeinschaft Steirisches Kürbiskernöl g.g.A.“ zur Verfügung gestellt (CRETNIK 2013, s.p.). Dies ist zwar keine vollständige Liste der insgesamt rund 70 österreichischen Ölmühlen, die Kürbiskernöl verarbeiten (die genaue Anzahl ist nicht bekannt), sie umfasst jedoch die größten Ölmühlen (CRETNIK 2013, s.p.). Beim Rest handelt es sich um sehr kleine Ölmühlen (CRETNIK 2013, s.p.). Zusätzlich sind noch die Standorte von 5 Wasch- bzw. Trocknungsanlagen im nördlichen Niederösterreich, ebenfalls als Punkte, eingezeichnet (MASCHINENRING 2010, 12). Auch diese Informationen sind leider lückenhaft.

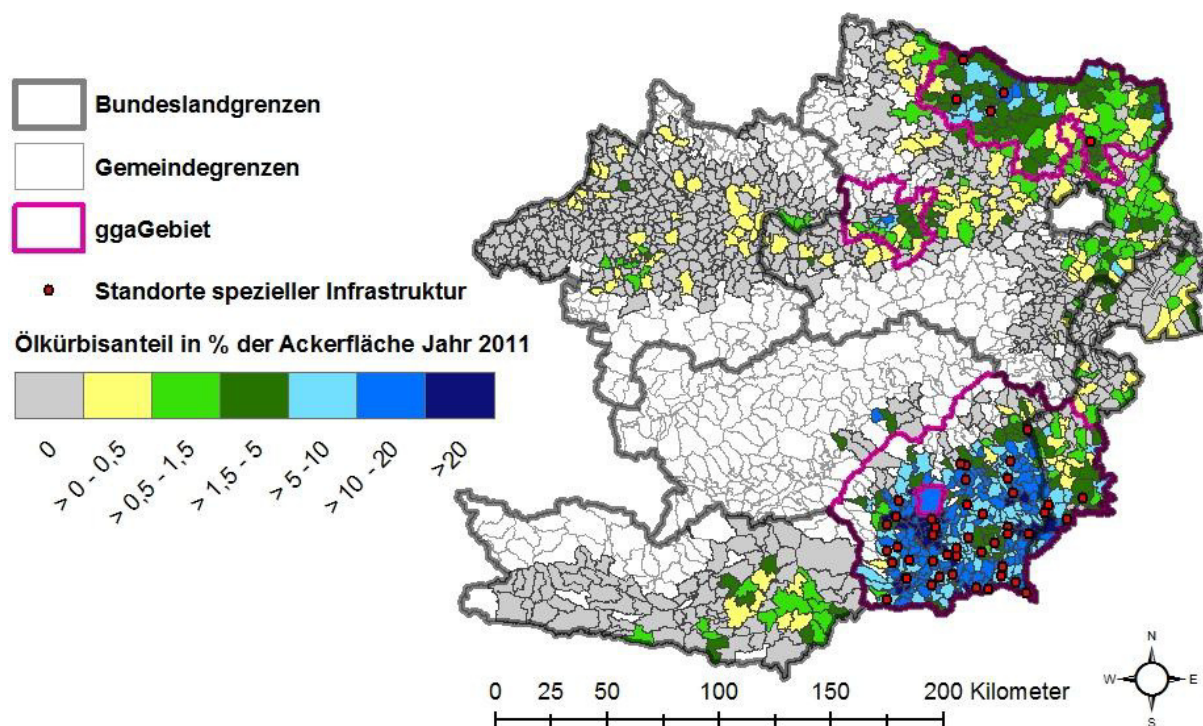
In den Abbildungen ist ersichtlich, dass der Schwerpunkt des Ölkürbisanbaus im g.g.A.-Gebiet liegt. Die Möglichkeit, die Produktion nach den Auflagen des Herkunftsschutzes durchzuführen, scheint sich positiv auf den Anbau auszuwirken. Neben den höheren

Erzeugerpreisen könnten die bereits vorhandene spezielle Infrastruktur und Know-How den Anbau innerhalb des g.g.A. Gebietes begünstigen. Speziell im g.g.A.-Gebiet in Niederösterreich wurde der Ölkürbisanbau in diesem Zeitraum durch Vertragsanbaufirmen stark forciert (BRANDSTETTER 2013, s.p.). Außerhalb des g.g.A.-Gebietes wird Ölkürbis in geringerem Ausmaß auch im Marchfeld, dem Kärntner Becken und Teilen Oberösterreichs angebaut. Die Anbauflächen nahmen zwischen 2000 und 2011 innerhalb des g.g.A.-Gebietes im nördlichen Weinviertel, der Steiermark sowie dem südlichen Burgenland und außerhalb des g.g.A. Gebietes im Marchfeld am stärksten zu. Der Großteil der Anbauflächen im Marchfeld dient zur Saatgutvermehrung, wofür kein Anbau im g.g.A.-Gebiet für einen erzielbaren Preisvorteil notwendig ist (BRANDSTETTER 2013, s.p.).



**Abbildung 4: Räumliche Verteilung des Ölkürbisianbaus in Oberösterreich, Niederösterreich, Burgenland, Steiermark und Kärnten 2000 (N = 1347 Gemeinden)**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)



**Abbildung 5: Räumliche Verteilung des Ölkürbisianbaus in Oberösterreich, Niederösterreich, Burgenland, Steiermark und Kärnten 2011 (N = 1347 Gemeinden)**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)



### **3 Material und Methoden**

Ökonometrische Modelle werden verwendet, um anhand empirischer Daten ökonomische Theorien auf ihre Richtigkeit zu testen, ökonomische Phänomene zu quantifizieren oder gesellschafts- bzw. wirtschaftspolitische Maßnahmen zu evaluieren. Die Regressionsanalyse stellt in diesem Zusammenhang eines der am häufigsten verwendeten Analyseverfahren dar (WOOLDRIDGE 2009, 1).

In diesem Kapitel werden Regressionsmodelle vorgestellt, mit denen sich die Bestimmungsfaktoren des Ölkürbisbaus in Österreich untersuchen lassen. Zunächst wird der räumliche und zeitliche Untersuchungsbereich der Analyse abgegrenzt. Danach werden die ermittelten Bestimmungsfaktoren des Ölkürbisbaus sowie deren Datenquellen beschrieben. Im Anschluss daran findet eine Erläuterung der verwendeten Regressionsmodelle und statistischen Tests statt. Zum Abschluss wird auf die Bildung der Modellgleichung sowie Interpretation der Regressionskoeffizienten eingegangen.

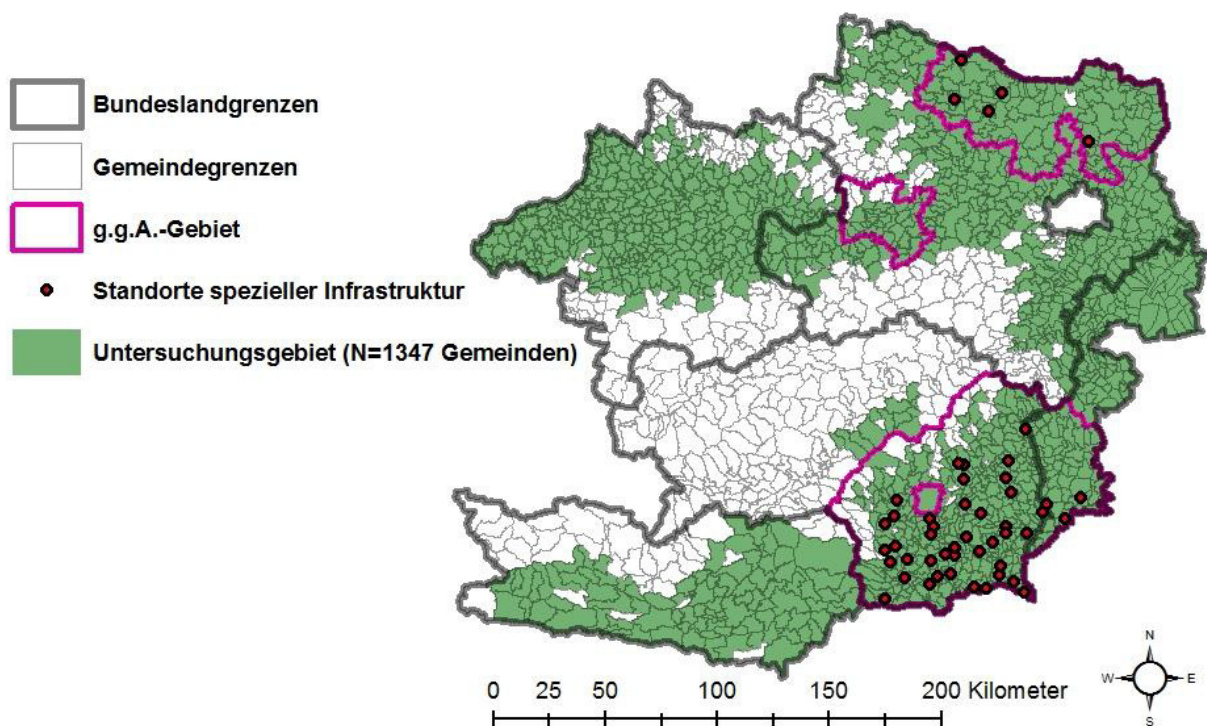
#### **3.1 Untersuchungsbereich**

Der räumliche Untersuchungsbereich wird auf die Bundesländer eingegrenzt, in denen erwähnenswerter Ölkürbisbau stattfindet. Dies sind: Steiermark, Niederösterreich, Burgenland, Kärnten und Oberösterreich.

Als räumliche Untersuchungsebene der Daten wird die Gemeindeebene gewählt. Ein Problem bei dieser Betrachtungsebene ist, dass die Effekte, welche die Ölkürbisbaufläche beeinflussen (darunter auch etwaige räumliche Effekte), auf einzelbetrieblicher Ebene wirken: Die Landwirte und Landwirtinnen entscheiden, ob und wie viel Ölkürbis sie anbauen. Betrachtet man die Daten auf Gemeindeebene sind nur räumliche Effekte beobachtbar, welche die Gemeindegrenzen überschreiten. Räumliche Effekte innerhalb einer Gemeinde entziehen sich der Wahrnehmung. Das bedeutet, die Ebene auf der die Entscheidungsträger (Landwirte bzw. Landwirtinnen) über den Anbau einer Kultur entscheiden, unterscheidet sich von der Betrachtungsebene der Regressionsanalyse (Gemeindeebene). In der Literatur wird dieser Effekt als „Modifiable Areal Unit Problem“ (MAUP) bezeichnet (ANSELIN 1988, 26). ANSELIN (2002, 261) weist in diesem Zusammenhang auf einen „Cross-Level Bias“ bei der „Ecological Regression“ hin, wodurch aufgrund der Aggregation der Variablen ungültige Rückschlüsse aus den Ergebnissen gezogen werden können. Bei Wahl der einzelbetrieblichen Ebene (Ebene der Entscheidungsträger) als räumliche Untersuchungsebene, würde man dieses Problem lösen, allerdings sind Daten auf einzelbetrieblicher Ebene nur begrenzt verfügbar.

Von den insgesamt 2.379 österreichischen Gemeinden (inklusive der Wiener Gemeindebezirke), die in der GEDABA gelistet sind (BMLFUW, 2012, s.p.), verbleiben nach Abzug der Gemeinden der Bundesländer Salzburg, Tirol, Vorarlberg und Wien 1.862 Gemeinden. Darin sind jedoch noch zahlreiche Gemeinden gelistet, in denen der Ölkürbisanbau aufgrund der klimatischen Bedingungen nicht oder nur bedingt möglich ist. Um diesen Umstand zu berücksichtigen, werden jene Gemeinden ausgeschieden, in denen im Zeitraum von 2000 bis 2011 nie Ölkürbis oder Körnermais angebaut wurde. Der Auswahlprozess wird anhand dieser Kriterien durchgeführt, da die klimatischen Ansprüche des Körnermais mit denen des Ölkürbis vergleichbar sind und der Körnermais im Vergleich zum Ölkürbis im österreichischen Ackerbau weiter verbreitet ist.

Einige Gemeinden können darüber hinaus nicht in die Analyse aufgenommen werden, da für sie nur lückenhafte bzw. inkompatible Daten verfügbar sind oder sie keine direkt angrenzende Nachbargemeinde haben. Dies ist von Bedeutung, da Gemeinden ohne Nachbarn (Inselbeobachtungen) bei der Berechnung eines räumlichen Regressionsmodells nicht zulässig sind. Nach diesen 3 Schritten verbleiben schließlich 1.347 Gemeinden als Untersuchungsgebiet im Modelldatensatz. Eine Übersicht dieser Gemeinden findet sich in Abbildung 6.



**Abbildung 6: Lage der Gemeinden des Untersuchungsgebietes**

*Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)*

Für den zeitlichen Untersuchungsbereich stehen, mit geringfügigen Einschränkungen, Daten von 2000 bis 2011 zur Verfügung. Einige mögliche Bestimmungsfaktoren sind jedoch über die Zeit hinweg konstant bzw. es wird unterstellt, dass sie sich im Analysezeitraum nicht verändert haben (etwa die Lage im g.g.A.-Gebiet sowie die klimatischen Bedingungen).

In vielen Gemeinden wird Ölkürbis nicht oder nur gelegentlich angebaut. Dieser Umstand engt die anwendbaren statistischen Verfahren zur Untersuchung der Bestimmungsfaktoren des Ölkürbisbaus auf Modelle mit einer beschränkten abhängigen Variable ein (siehe z. B. WOOLDRIDGE 2009, 588). Da grundsätzlich Daten für den Zeitraum von 2000 bis 2011 verfügbar sind, spricht das für die Anwendung eines Panelmodells. In der Literatur finden sich zwar bereits Anwendungsbeispiele räumlicher Panelmodelle mit vergleichbarer Datenstruktur, diese sind jedoch bislang vorwiegend auf eine methodologische Ebene beschränkt (siehe z. B. AUTANT-BERNARD und LESAGE 2009). Zusätzlich stellt der Mangel an verfügbaren Softwarepaketen für solche Problemstellungen ein zentrales Problem dar. Um sich auf den Fokus dieser Arbeit, die Untersuchung räumlicher Effekte mit Hilfe von räumlichen Regressionsmodellen, zu konzentrieren und um auch die zeitlich konstanten Variablen in der Analyse berücksichtigen zu können, werden die Berechnungen mit Querschnittsdaten für das aktuellste verfügbare Jahr (2011) durchgeführt.

### **3.2 Modellvariablen**

Der Ölkürbisbau wird durch viele Faktoren beeinflusst. Neben den klimatischen Voraussetzungen, könnten vor allem die Lage im g.g.A.-Gebiet und die Nähe zu spezieller Infrastruktur die räumliche Verteilung des Ölkürbisbaus beeinflussen. Für die Auswahl potentieller Bestimmungsfaktoren werden die Informationen aus Kapitel 2 herangezogen. Um die Daten für die Bestimmungsfaktoren in der Untersuchung verwenden zu können, müssen diese zuerst aufbereitet und teilweise (z.B. durch Logarithmierung) transformiert werden. Das ist aufgrund der teilweise sehr unregelmäßigen Verteilung der Variablen erforderlich.

#### **3.2.1 Abhängige Variable**

Der Ausgangspunkt für die Zielgröße der Regressionsanalyse ist der Anteil der Ölkürbisflächen an der gesamten Ackerfläche pro Gemeinde in Prozent (0-100). Es werden relative (Prozent), anstatt absoluter (Hektar) Werte für die abhängige Variable gewählt, um Verzerrungen aufgrund unterschiedlicher Gemeindegrößen zu vermeiden. In anderen Studien wird ebenfalls auf diese Weise vorgegangen (siehe z. B. SCHOLZ et al. 2013). Da die Variable Ölkürbisanteil eine sehr ungleichmäßige Verteilung aufweist, wurde sie logarithmiert ( $\ln(1 + OelkAnt)$ ). Die Konstante 1 wurde aufgrund der zahlreichen Nullen im Datensatz

addiert. Die Daten für den Ölkürbisanteil stammen aus der Gemeindedatenbank des BMLFUW (2012).

Die möglichen Bestimmungsfaktoren basieren auf Literaturangaben und Informationen von den im Rahmen dieser Arbeit befragten Experten. Ein computerunterstützter Auswahlprozess ergänzt die Festlegung der unabhängigen Variablen. Das Akaike-Information-Criterion (AIC) ist die Entscheidungsgröße für die Variablenauswahl (AKAIKE 1973). Das AIC beschreibt die Güte (Fit) eines Regressionsmodelles, indem es die „Distanz“ zwischen einem unbekanntem Mechanismus, der die beobachteten Daten erzeugt haben könnte, und einem mit den Daten geschätzten Modell wiedergibt (SACHS und HEDDERICH 2009, 664). Dem AIC wird ein Strafterm ( $2^x$  Anzahl der zu schätzenden Parameter) hinzugefügt, um eine Balance zwischen zu wenigen (Underfit) und zu vielen (Overfit) Einflussgrößen in einem Modell zu finden (SACHS und HEDDERICH 2009, 664). Je kleiner der AIC-Wert eines Modells, desto besser dessen Fit (SACHS und HEDDERICH 2009, 665).

Die Ergebnisse des Variablenauswahlprozesses auf Basis des AIC müssen jedoch zusätzlich einer sachlogischen Beurteilung unterzogen werden, um zu überprüfen, ob die so ermittelte Variablenkombination auch aus theoretischer Sicht sinnvoll erscheint. Die so ausgewählten Bestimmungsfaktoren, welche für die Regressionsanalyse Verwendung finden, werden nachfolgend vorgestellt. In weiterer Folge werden für die Bestimmungsfaktoren Hypothesen über die Art ihrer Wirkung auf die abhängige Variable (positiv oder negativ) gebildet.

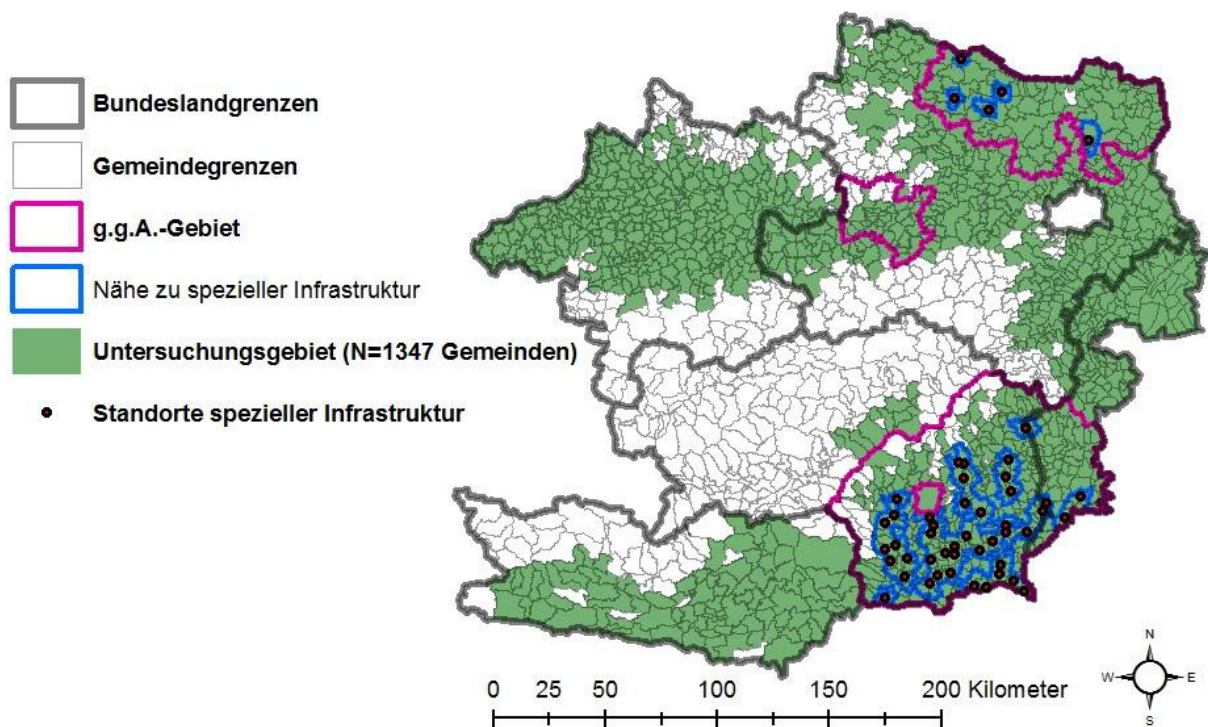
Mit den angeführten Bestimmungsfaktoren wird jedoch kein Anspruch auf Vollständigkeit erhoben, da für einige potentielle Bestimmungsfaktoren kein geeignetes bzw. nur unvollständiges Datenmaterial verfügbar ist, sodass sie bei der Analyse nicht, bzw. nur in eingeschränkter Form, mit einbezogen werden können.

### **3.2.2 Unabhängige Variablen**

- **Lage im g.g.A.-Gebiet und Nähe zu spezieller Infrastruktur:**

Die Lage im „Steirisches Kürbiskernöl g.g.A.“-Gebiet entscheidet darüber, ob das aus den Kürbiskernen gewonnene Öl als „Steirisches Kürbiskernöl g.g.A.“ verkauft werden kann (SCHWARZ 2008, 81). Darüber hinaus wird davon ausgegangen, dass die räumliche Nähe zu be- und verarbeitenden Betrieben (Ölmühlen sowie Wasch- bzw. Trocknungsanlagen) einen positiven Effekt auf den Ölkürbisanteil an der Ackerfläche innerhalb einer Gemeinde hat. Diese Verarbeitungsschritte sind auf einzelbetrieblicher Ebene meist nicht wirtschaftlich. Es werden daher 2 Dummy-Variablen („Lage der Gemeinde im g.g.A.-Gebiet“ und „Nähe zu spezieller Infrastruktur“) gebildet, um diese Wirkungen im Modell zu berücksichtigen. Die

Variable „Lage der Gemeinde im g.g.A.-Gebiet“ kann die drei Ausprägungen „Gemeinde liegt nicht im g.g.A.-Gebiet“ (838 Gemeinden), „Gemeinde liegt im g.g.A.-Gebiet in Niederösterreich“ (134 Gemeinden) und „Gemeinde liegt im g.g.A.-Gebiet in der Steiermark bzw. dem südlichen Burgenland“ (375 Gemeinden) annehmen. Es wird davon ausgegangen, dass die Lage im g.g.A.-Gebiet einen positiven Effekt auf den Ölkürbisanteil der Gemeinden hat. Die Unterteilung des g.g.A.-Gebietes in 2 Teile soll den unterschiedlichen Produktions- und Vermarktungsstrukturen der beiden Regionen Rechnung tragen. Mit der zweiten Dummy-Variable wird die Nähe zu spezieller Infrastruktur (Ölmühle, Wasch- bzw. Trocknungsanlage) beschrieben. Als in der Nähe von spezieller Infrastruktur werden Gemeinden definiert, deren Mittelpunkt innerhalb eines 5 km-Radius zu einem Mittelpunkt einer Gemeinde mit spezieller Infrastruktur liegt. In Abbildung 7 sind die räumlichen Ausprägungen der beiden Dummy-Variablen grafisch dargestellt.



**Abbildung 7: Räumliche Darstellung der Dummy Variablen**

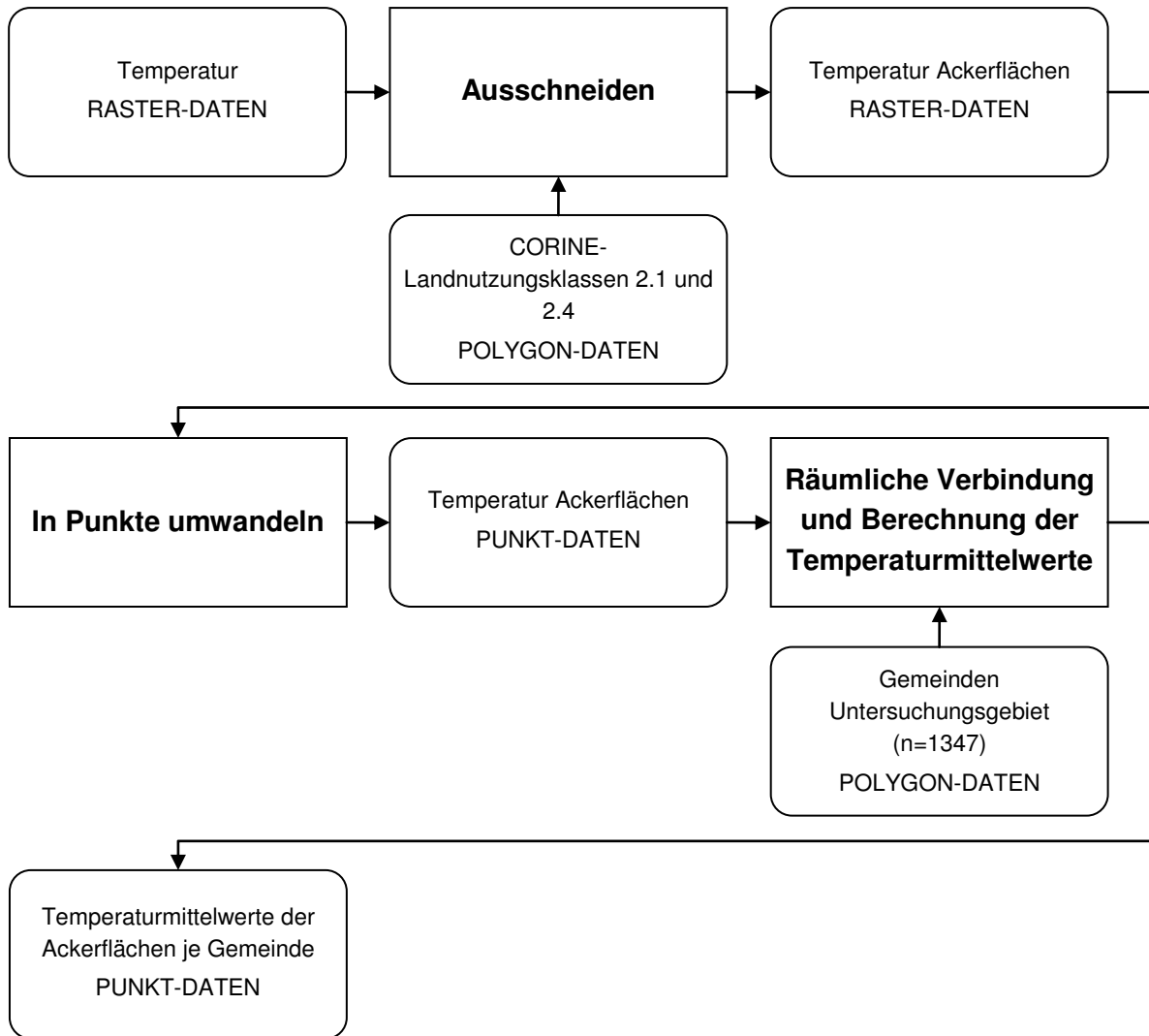
Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)

• **Temperatur:**

Um die Wärmebedürftigkeit des Ölkürbis zu berücksichtigen, werden die Jahresdurchschnittstemperatur in °C und das Quadrat der Jahresdurchschnittstemperatur (Temp<sup>2</sup>) als weitere unabhängige Variablen gewählt. Es wird dabei von einem positiven Zusammenhang zwischen der Jahresdurchschnittstemperatur und dem Anteil der Ölkürbisfläche an der Ackerfläche ausgegangen. Die quadrierte Variable wird in das Modell

aufgenommen, da sie einerseits den Modell-Fit erhöht und andererseits aus pflanzenbaulicher Sicht von einem abnehmenden positiven Temperatureffekt mit steigender Temperatur ausgegangen werden kann. Die Durchschnittstemperatur für den Vegetationszeitraum des Ölkürbis (April-September) wäre als Maßzahl besser geeignet, diese Daten stehen aber nicht zur Verfügung.

Die Datengrundlage stellen langjährige Temperaturmittelwerte (30-jährige Mittelwerte über den Zeitraum von 1971-2000) dar (ZAMG 2013, s.p.). Die Daten, die in Form von Rasterdaten mit einer Auflösung von 250 m x 250 m vorliegen, müssen zunächst mit Hilfe der Software ArcGIS (ESRI 2011) auf Gemeindeebene aggregiert werden, bevor sie für die Analyse verwendet werden können. Bei einer simplen Aggregation der Jahresdurchschnittstemperaturen auf Gemeindeebene, würden jedoch speziell in Gemeinden mit großen Höhenunterschieden starke Temperaturverzerrungen auftreten. Befindet sich etwa innerhalb einer Gemeinde eine Rasterzelle im Tal mit einem Temperaturwert von 8,2 °C und eine Rasterzelle in 2000m Höhe mit einem Wert von 1,5 °C, so würde der daraus gebildete Mittelwert jegliche Aussagekraft für die Analyse verlieren. Um diese Verzerrungen zu vermeiden, wird versucht, nur die Temperaturwerte von Ackerflächen zu aggregieren. Die dafür notwendigen räumlichen Informationen über die Landnutzung liefern die CORINE-Landnutzungsdaten (Coordination of Information on the Environment) (UMWELTBUNDESAMT 2006, s.p.). Basierend auf diesen Informationen erfolgt zuerst eine Verschneidung der Temperaturresterdaten mit den CORINE-Landnutzungsklassen 2.1 (Ackerflächen) und 2.4 (Flächen mit heterogener Struktur). Danach werden die so bereinigten Temperaturdaten auf Gemeindeebene aggregiert. Das Resultat daraus ist die mittlere Temperatur der Ackerflächen je Gemeinde. Die dafür in ArcGIS durchgeführten Berechnungen sind in Abbildung 8 schematisch dargestellt. Die Felder mit abgerundeten Ecken repräsentieren Input- und Outputdaten und die rechteckigen Felder beschreiben Operationen in ArcGIS.



**Abbildung 8: Flow-Chart der Aufbereitung der Temperaturdaten**

• **Niederschlag:**

Als weitere Klimavariablen werden die durchschnittlichen Niederschläge von April bis September in l/m<sup>2</sup> herangezogen. Da ausreichend Niederschläge für gute Kernerträge notwendig sind, wird von einem positiven Effekt des Niederschlags auf den Ölkürbisanteil der Gemeinden ausgegangen. Für die tatsächliche Wirkung der Niederschläge ist jedoch die genaue zeitliche Verteilung der Niederschlagsmengen entscheidender, dafür stehen jedoch keine genaueren Daten zur Verfügung. Die Niederschlagsdaten stammen ebenfalls von der ZAMG und liegen in derselben Form wie die Temperaturdaten vor (ZAMG 2013, s.p.). Sie werden direkt, ohne Berücksichtigung der CORINE-Landnutzungsklassen, auf Gemeindeebene aggregiert, da die Niederschläge im Vergleich zur Temperatur weitaus weniger kleinräumigen Schwankungen unterliegen.

- **Tierhaltung:**

Die produktionstechnische Ausrichtung eines landwirtschaftlichen Betriebes nimmt großen Einfluss auf die Entscheidung, ob und welche Ackerkulturen angebaut werden. So erfordert die Tierhaltung meist einen höheren Arbeitsaufwand als der reine Ackerbau. Außerdem benötigen tierhaltende Betriebe ihre Ackerflächen auch für den Anbau von Futterpflanzen. Die Kultivierung von Ölkürbis erfordert im Vergleich zu vielen anderen Ackerkulturen einen höheren Arbeitsaufwand. Es wird daher davon ausgegangen, dass in Gemeinden mit viel Tierhaltung weniger Ölkürbis angebaut. Mit Hilfe der Anzahl der Großvieheinheiten (GVE) pro ha landwirtschaftlich genutzte Fläche (LF) je Gemeinde soll diese Hypothese überprüft werden. Die Daten für die GVE pro ha LF sowie für die nachfolgenden Bestimmungsfaktoren Ackerfläche, LF pro Betrieb und Körnermaisanteil stammen aus der GEDABA (BMLFUW 2012).

- **Körnermaisanteil:**

Da sich die klimatischen Ansprüche des Ölkürbis und des Körnermais sehr ähnlich sind, kann aus einem hohen Körnermaisanteil, bezogen auf die Ackerfläche, indirekt auch auf einen hohen Ölkürbisanteil geschlossen werden. Dabei gilt es jedoch zu beachten, dass ab einem gewissen Prozentsatz die Ausdehnung der einen Ackerkultur zu Lasten der anderen geht. Um diese Effekte beschreiben zu können werden 2 Proxyvariablen für die klimatische Eignung des Ölkürbisbaus berechnet: Der Anteil der Körnermaisflächen an der gesamten Ackerfläche je Gemeinde sowie das Quadrat des Körnermaisanteils ( $KmAnt^2$ ). Die Annahme ist, dass sich ein höherer Körnermaisanteil positiv auf den Ölkürbisanteil auswirkt. Mit zunehmendem Körnermaisanteil ergibt sich jedoch eine Konkurrenzsituation, wodurch ein negativer Effekt entsteht. Dieser funktionale Zusammenhang lässt sich mit einer quadrierten Variable im Modell darstellen. Da die Körnermaisanteile sehr ungleichmäßig verteilt sind, werden die beiden Variablen ebenfalls für die Regressionsanalyse logarithmiert. Die quadrierten Körnermaisanteile werden dabei auf Basis der logarithmierten Körnermaisanteile berechnet.

- **Ackerfläche:**

Eine Voraussetzung für den Ölkürbisbau sind vorhandene Ackerflächen. Daher wird als weitere Variable die Ackerfläche in ha je Gemeinde gewählt. Mit dieser Variable lässt sich kontrollieren, ob sich der Ölkürbisanteil mit zunehmender Ackerfläche verändert. Der Ölkürbisanteil einer Gemeinde ist als „Ölkürbisfläche / Ackerfläche“ definiert. Daraus folgt, dass der Ölkürbisanteil entweder durch eine Veränderung der Ölkürbisfläche und/oder der Ackerfläche beeinflusst wird. Durch die Berücksichtigung der Ackerfläche als unabhängige Variable, lässt sich der Einfluss der anderen Bestimmungsfaktoren auf den Ölkürbisanteil bei



konstant gehaltener Ackerfläche schätzen. Es wird davon ausgegangen, dass eine größere Summe der Ackerflächen sich positiv auf den Ölkürbisanteil einer Gemeinde auswirkt. Für die Regressionsanalyse wird die Variable aufgrund ihrer ungleichmäßigen Verteilung logarithmiert.

- **Landwirtschaftlich genutzte Fläche pro Betrieb:**

Auch die Größenstruktur der landwirtschaftlichen Betriebe könnte den Ölkürbisanteil einer Gemeinde beeinflussen. Während große Betriebe Skaleneffekte nutzen können, wodurch z.B. mechanisierte Produktionsverfahren rentabler werden, sind kleine Betriebe darauf angewiesen, eine höhere Wertschöpfung pro Flächeneinheit zu erzielen, um ihr Einkommen aus der Landwirtschaft zu sichern. Kleine Betriebe können das z.B. mit der Direktvermarktung von Kürbiskernöl erreichen. Es wird daher davon ausgegangen, dass in Gemeinden mit einer höheren durchschnittlichen Betriebsgröße der Ölkürbisanteil tendenziell niedriger ist. Die Größenstruktur der Betriebe wird durch die durchschnittliche landwirtschaftlich genutzte Fläche in ha pro Betrieb beschrieben. Um den Effekt von Ausreißern (Gemeinden mit überdurchschnittlich hoher LF pro Betrieb) zu mindern, wird diese Variable ebenfalls logarithmiert.

- **Räumliche Abhängigkeit:**

Der letzte potentielle Bestimmungsfaktor beruht auf der Annahme, dass der Ölkürbisanteil einer Gemeinde durch Faktoren in benachbarten Gemeinden beeinflusst wird. Räumliche Abhängigkeit könnte z. B. auf der Nähe zu nicht erfassten, be- und verarbeitenden Betrieben (Wasch- und Trocknungsanlagen bzw. Ölmühlen), überbetrieblichen Kooperationen für Anbau, Ernte, Waschen und Trocknen oder Erfahrungs- und Wissensaustausch zwischen landwirtschaftlichen Betrieben beruhen. Für einen Teil dieser Faktoren liegen zwar standortbezogene Daten vor, sie sind aber sehr lückenhaft. Durch die Erfassung räumlicher Effekte wird versucht, diese fehlenden Informationen zu berücksichtigen.

### **3.2.3 Deskriptive Statistik der Modellvariablen**

In Tabelle 1 sind neben einer Zusammenfassung der Hypothesen zu den einzelnen Variablen einige aussagekräftige Maßzahlen der deskriptiven Statistik enthalten. Minimum und Maximum zeigen den Wertebereich innerhalb dessen sich die Variablen bewegen. Erstes und drittes Quartil sowie der Mittelwert beschreiben die Verteilung der Werte im Wertebereich. Für die Dummy Variablen finden sich neben den möglichen Ausprägungen in Klammer die Anzahl der Gemeinden. Besonders auffällig ist, dass der Log-transformierte Ölkürbisanteil in 702 Gemeinden (ca. 52%) bei 0 liegt, was für die Verwendung eines Tobit-Modelles spricht.

Von den 1347 Gemeinden liegen insgesamt 134 im g.g.A.-Gebiet in Niederösterreich und 375 im g.g.A.-Gebiet in der Steiermark und dem Burgenland. Der größte Anteil der Gemeinden (838) liegt außerhalb des g.g.A.-Gebietes. Insgesamt 170 Gemeinden befinden sich in der Nähe von spezieller Infrastruktur. Die Jahresdurchschnittstemperatur bewegt sich zwischen 5,8 und 10°C. Da das Untersuchungsgebiet auf Basis klimatischer Bedingungen eingeschränkt wurde, sind viele Gemeinden mit niedrigerer Jahresdurchschnittstemperatur weggefallen. Durch die Logarithmus-Transformation liegen die letzten 4 Bestimmungsfaktoren näher an einer Normalverteilung und der Effekt von Ausreißern ist geringer.

**Tabelle 1: Zusammenfassung der Modellvariablen – Hypothesen und deskriptive Statistik (n=1347 Gemeinden)**

| Merkmal<br>Modellvariablen                           | Hypothese<br>(+, +/-, -) | Minimum  | Erstes<br>Quartil | Mittelwert | Drittes<br>Quartil | Maximum |
|--|--------------------------|----------|-------------------|------------|--------------------|---------|
| Ölkürbisanteil<br>(Log-transformiert)                | NA*                      | 0,0      | 0,0               | 0,7        | 1,4                | 3,9     |
| Dummy g.g.A.-<br>Gebiet                              | +                        | 0 (838)  | NA                | 1 (134)    | NA                 | 2 (375) |
| Dummy spezielle<br>Infrastruktur                     | +                        | 0 (1177) | NA                | NA         | NA                 | 1 (170) |
| Temperatur in °C                                     | +                        | 5,8      | 8,0               | 8,5        | 9,1                | 10,0    |
| (Temperatur) <sup>2</sup>                            | -                        | 33,8     | 63,4              | 73,3       | 82,8               | 100,5   |
| Niederschlag in l/m <sup>2</sup>                     | +/-                      | 310,5    | 416,3             | 509,2      | 587,9              | 942,7   |
| GVE pro ha LF  | -                        | 0,0      | 0,2               | 0,7        | 1,1                | 2,1     |
| Körnermaisanteil<br>(Log-transformiert)              | +                        | 0,0      | 1,4               | 2,3        | 3,3                | 4,4     |
| Körnermaisanteil <sup>2</sup><br>(Log-transformiert) | -                        | 0,0      | 2,0               | 6,8        | 11,0               | 19,7    |
| Ackerfläche<br>(Log-transformiert)                   | +                        | 0,0      | 5,6               | 6,2        | 7,1                | 9,1     |
| LF pro Betrieb<br>(Log-transformiert)                | -                        | 0,8      | 2,6               | 3,0        | 3,4                | 6,8     |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013

\*: Felder, die mit NA gekennzeichnet sind haben keine Merkmalsausprägung

### 3.3 Ökonometrische Modelle und statistische Tests

In Abschnitt 3.3.1 wird ein einfaches lineares Regressionsmodell (OLS-Modell) aufgestellt und seine Funktionsweise erläutert. Darüber hinaus werden die verwendeten statistischen Tests zur Überprüfung der Modellannahmen, der Reset-Test und der Breusch-Pagan-Test,

vorgestellt. Die hohe Anzahl an Gemeinden im Datensatz, in denen kein Ölkürbis angebaut wird, hat zur Folge, dass in einem Modell nach der Methode der kleinsten Quadrate inkonsistente Schätzergebnisse zu erwarten sind (LONG 1997, 189). Daher wird in Abschnitt 3.3.2 ein Tobit-Modell gebildet, das speziell für solche Fälle konzipiert wurde (WOOLDRIDGE 2009, 588). Ein weiterer Aspekt, der die Ergebnisse der Modelle verzerren kann, ist räumliche Autokorrelation (ANSELIN 1988, 58ff). In Abschnitt 3.3.3 wird zunächst ein allgemeiner Überblick über das Konzept und die Auswirkungen von räumlicher Autokorrelation gegeben und statistische Testverfahren vorgestellt, mit denen räumliche Autokorrelation festgestellt werden kann. Abschließend wird die Berücksichtigung von räumlicher Autokorrelation im linearen Regressionsmodell und im Tobit-Modell erläutert.

Nachfolgend werden nur die wesentlichsten Punkte sowie notwendige Annahmen für die einzelnen Modelle besprochen. Für detailliertere Informationen zu den jeweiligen Modellen, den angewendeten Schätzverfahren und den statistischen Tests siehe etwa BACKHAUS et al. (2011), FAHRMEIR et al. (2009) WOOLDRIDGE (2009), LONG (1997), PLANT (2012) oder LESAGE und PACE (2009).

### 3.3.1 Lineares Regressionsmodell

Die Grundidee der linearen Regressionsanalyse besteht in der Modellierung der Ursache-Wirkungs-Beziehung (Kausalbeziehung) von einer oder mehrerer unabhängigen Variable(n)  $X$ , die auf eine abhängige Variable  $Y$  wirken, mit Hilfe eines linearen Gleichungssystems (BACKHAUS et al. 2011, 56). Diese Beziehung kann in einer allgemeinen Formel in Matrixform folgendermaßen dargestellt werden (WOOLDRIDGE 2009, 22):

$$Y = \alpha + \beta X + U$$

mit

$Y$  = Vektor der abhängigen Variable mit  $n$  Beobachtungen

$X$  = Matrix ( $n \times (k+1)$ ) mit den Ausprägungen der unabhängigen Variablen ( $k$  = Anzahl der unabhängigen Variablen)

$\alpha$  = konstanter Term

$\beta$  = Vektor der  $k$  Regressionskoeffizienten

$U$  = Vektor der  $n$  Fehler

In der Praxis sind jedoch nur die Werte der abhängigen Variable und der unabhängigen Variablen bekannt. Die Regressionskoeffizienten, welche die Stärke des Einflusses einer unabhängigen Variable auf die abhängige Variable beschreiben, müssen mit Hilfe eines Verfahrens geschätzt werden. Die Schätzung der Parameter führt zu folgender Regressionsgleichung (WOOLDRIDGE 2009, 32):

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X$$

mit

$\hat{Y}$  = Vektor der Schätzer der abhängigen Variable (gefittete Werte) mit  $n$  Ausprägungen

$\hat{\alpha}$  = Schätzer des konstanten Terms

$X$  = Matrix ( $n * (k+1)$ ) mit den Ausprägungen der unabhängigen Variablen

$\hat{\beta}$  = Vektor der Schätzer der  $k$  Regressionskoeffizienten

Die Regressionsgleichung stellt eine Schätzung der tatsächlichen Beziehung dar. Daher weichen die geschätzten (gefitteten) Werte von den tatsächlichen Ausprägungen der abhängigen Variable ab. Die Differenz aus beobachteten und gefitteten Werten ergibt die Schätzwerte der Fehler (Residuen) (WOOLDRIDGE 2009, 31):

$$\hat{U} = Y - \hat{Y}$$

mit

$\hat{U}$  = Vektor der Schätzer der Fehler (Residuen) mit  $n$  Ausprägungen

Bei der Schätzung nach der Methode der kleinsten Quadrate (in Englisch „Ordinary Least Square“, kurz OLS) werden die unbekannt Parameter so bestimmt, dass die Summe der quadrierten Residuen minimal wird (BACKHAUS et al. 2011, 67):

$$\hat{U} = [Y - (\hat{\alpha} + \hat{\beta}X)]^2 \rightarrow \min$$

Die Residuen werden in quadrierter Form minimiert, um einerseits zu verhindern, dass sich positive und negative Abweichungen gegenseitig aufheben und andererseits größere Abweichungen stärker zu gewichten (BACKHAUS et al. 2011, 67).

Die Schätzer der Methode der kleinsten Quadrate sind unter folgenden 5 Annahmen BLUE („Best Linear Unbiased Estimators“) (WOOLDRIDGE, 2009, 103ff):

- Das Modell ist linear in seinen Parametern
- Die X- und Y-Werte sind Ergebnis einer Zufallsstichprobe oder stellen die Gesamtpopulation dar
- Die unabhängigen Variablen sind keine linearen Kombinationen voneinander
- Der Fehlerterm  $u$  hat einen Erwartungswert von 0 und ist somit unabhängig von den Ausprägungen der X-Werte
- Die Varianz des Fehlerterms ist unabhängig von den Ausprägungen der X-Werte (Homoskedastizität)

Während die ersten 3 Annahmen sehr allgemeiner Natur sind, bedarf die Überprüfung der letzten beiden Annahmen statistischer Tests (ECKEY et al. 2011, 100f).

Ein wesentlicher Aspekt bei statistischen Tests ist das Signifikanzniveau, auf dem die Entscheidung getroffen wird, ob ein Ergebnis als signifikant erachtet wird. Im Rahmen dieser Arbeit wird für statistische Tests ein Signifikanzniveau von 5% verwendet, das im wirtschafts- und sozialwissenschaftlichen Bereich üblicherweise angewandt wird (WOOLDRIDGE 2009, 123).

Bei Verletzung der vierten Annahme (Erwartungswert des Fehlerterms = 0) sind die Schätzer der Regressionskoeffizienten  $\hat{\beta}$  verzerrt. Dies kann etwa auf eine falsche Modellierung des funktionalen Zusammenhangs oder das Weglassen einer wichtigen unabhängigen Variable zurückführbar sein (Fehlspezifikation) (WOOLDRIDGE 2009, 87ff). Um diese Annahme zu überprüfen, kann der RESET-Test angewendet werden (ECKEY et al. 2011, 224). Ist das Ergebnis des Tests auf dem 5%-Niveau signifikant, so wird in der vorliegenden Arbeit die Nullhypothese (keine Fehlspezifikation) verworfen.

Wird die Annahme von Homoskedastizität verletzt, so liegt Heteroskedastizität (Varianzinhomogenität) vor, was die Varianz der Schätzer der Regressionskoeffizienten  $\hat{\beta}$  verzerrt (WOOLDRIDGE 2009, 265). Dadurch verlieren Signifikanztests und Konfidenzintervalle der Schätzer ihre Gültigkeit (WOOLDRIDGE 2009, 265). Die Annahme von Homoskedastizität lässt sich mit dem Breusch-Pagan-Test überprüfen. Im Breusch-Pagan-Test wird von der Nullhypothese Homoskedastizität ausgegangen, die gegen die Alternativhypothese Heteroskedastizität getestet wird (ECKEY et al. 2011, 109). Die Nullhypothese wird im Rahmen dieser Arbeit wiederum verworfen, wenn das Ergebnis auf dem 5% Niveau signifikant ist. Liegt Heteroskedastizität vor, so kann dieses Problem bei einem OLS-Modell mit der Berechnung heteroskedastischer Standardfehler der Schätzer  $\hat{\beta}$  gelöst werden (WOOLDRIDGE 2009, 265).

### **3.3.2 Tobit-Modell**

Das Tobit-Modell wird angewendet, wenn die Ausprägungen einer kontinuierlichen abhängigen Variable durch einen Grenzwert eingeschränkt sind und ein größerer Teil der abhängigen Variablen den Grenzwert annimmt (Long 1997, 187). In der Praxis liegt dieser Wert häufig bei 0. Aus den Daten im Rahmen dieser Arbeit geht hervor, dass in 702 der 1.347 untersuchten Gemeinden (ca. 52%) im Jahr 2011 kein Ölkürbis angebaut wurde. In einem solchen Fall führt eine OLS-Regression zu inkonsistenten Schätzern der Regressionsparameter (Long, 1997, 189). Im Tobit-Modell wird zur Lösung dieses Problems

eine latente Variable  $Y^*$  für die abhängige Variable  $Y$  herangezogen, die mit folgender Gleichung beschrieben werden kann (Long 1997, 196):

$$Y^* = \alpha + \beta X + U$$

mit

$Y^*$  = Vektor der latenten abhängigen Variablen mit  $n$  Ausprägungen

$\alpha$  = konstanter Term

$X$  = Matrix ( $n * (k+1)$ ) mit den Ausprägungen der unabhängigen Variablen

$\beta$  = Vektor der  $k$  Regressionskoeffizienten

$U$  = Vektor der  $n$  Fehler

Die einzelnen latenten Variablen ( $y_i^*$ ) stehen dabei mit den beobachteten abhängigen Variablen  $y_i$  in folgender Beziehung (Long, 1997, 196):

$$y_i = y_i^* \text{ wenn } y_i^* > 0$$

und

$$y_i = 0 \text{ wenn } y_i^* < 0$$

Das obige lineare Gleichungssystem beschreibt also einerseits den Einfluss der unabhängigen Variablen  $X$  auf die latente Variable  $Y^*$  sowie die Verknüpfung der Variablen  $Y^*$  und  $Y$ . Für die Variable  $Y$  gilt, dass sie gleich  $Y^*$  ist, sobald  $Y^*$  größer als 0 ist und 0, wenn  $Y^*$  kleiner als 0 ist (WOOLDRIDGE 2009, 588). In Tobit-Modellen wird als implizit unterstellt, dass die latente abhängige Variable  $Y^*$  kontinuierlich verteilt ist und gleichzeitig die abhängige Variable  $Y$  nur positive Werte annehmen kann (Wooldridge 2009, 588).

Die Regressionsparameter werden im Tobit-Modell meist iterativ mittels Maximum-Likelihood-Methode (ML) geschätzt (LONG 1997, 204): Dabei werden die Parameter unter der Annahme ermittelt, dass  $Y^*$  aus einer bestimmten Wahrscheinlichkeitsverteilung (meist Normalverteilung) stammt. Die Maximum-Likelihood-Schätzer werden durch Maximierung der Wahrscheinlichkeit, die erhobenen Daten zu erhalten, berechnet. Die Formel für ein Tobit-Modell mit einem Grenzwert von 0 und unter Annahme einer Normalverteilung von  $Y^*$  lautet (LONG, 1997, 205):

$$\ln L(\beta, \sigma^2 | Y, X) = \sum_{y>0} \ln \frac{1}{\sigma} \phi\left(\frac{y_i - x_{i,k} \beta_k}{\sigma}\right) + \sum_{y<0} \ln \Phi\left(\frac{0 - x_{i,k} \beta_k}{\sigma}\right) \rightarrow \max$$

mit

$\ln$  = natürlicher Logarithmus

$L$  = Likelihood-Funktion

$\beta$  = Vektor der  $k$  Regressionskoeffizienten

$\sigma$  = Standardabweichung der Fehler

$Y$  = Vektor der abhängigen Variablen mit  $n$  Ausprägungen

$X$  = Matrix ( $n * (k+1)$ ) mit den Ausprägungen der unabhängigen Variablen

$\phi$  = Dichtefunktion der Normalverteilung

$\Phi$  = kumulative Normalverteilungsfunktion

Zentrale Modellannahmen eines Tobit-Modells nach ML-Schätzung sind jedoch, dass die Fehler ( $U$ ) homoskedastisch sind und einer Normalverteilung folgen (LONG 1997, 206).

### 3.3.3 Räumliche Regressionsmodelle

Die räumliche Regressionsanalyse unterscheidet sich von einer herkömmlichen Regressionsanalyse dadurch, dass räumliche Abhängigkeit (räumliche Autokorrelation) mit in die Modellberechnungen einbezogen wird. Räumliche Abhängigkeit bedeutet, dass sich die Werte räumlich naher Beobachtungen in ihren Ausprägungen ähneln (PLANT 2012, 59). Unterstellt man diesem Phänomen einen funktionellen Zusammenhang, dann beeinflusst das was an einem Ort passiert, das was sich an einem benachbarten Ort ereignet (ANSELIN 1988, 11). Aus theoretischer Sicht kann räumliche Autokorrelation auf den 3 Phänomenen Interaktion, Reaktion und Fehlspezifikation beruhen (PLANT 2012, 423f):

Interaktion bedeutet, dass sich die abhängigen Variablen (Anteil der Ölkürbisflächen an der Ackerfläche) benachbarter Gemeinden gegenseitig beeinflussen. Dies könnte etwa auf einer Form von Wissensaustausch, gegenseitiger Beobachtung, bzw. gemeinsamer Nutzung von Ressourcen (überbetriebliche Korporationen) bei Landwirten bzw. Landwirtinnen beruhen. Wird diese Form von räumlicher Autokorrelation nicht berücksichtigt, hat das verzerrte Schätzer der Regressionsparameter zur Folge (KLOTZ 2004, 39ff).

Reaktion liegt vor, wenn die Ölkürbisanteile benachbarter Gemeinden auf externe Faktoren reagieren, welche im Raum variieren. Ein Beispiel hierfür wäre etwa die Entfernung zur nächsten Wasch- und Trocknungsanlage. Dabei wird davon ausgegangen, dass ein positiver Einfluss besteht, der mit steigender Entfernung abnimmt. Diese Form von räumlicher Autokorrelation kann durch die Aufnahme einer entsprechenden Variable (Distanz zur nächsten Wasch- und Trocknungsanlage) in das Modell berücksichtigt werden. Stehen keine Daten für die Variable zur Verfügung, kann das unterschiedliche Auswirkungen auf andere Modellvariablen oder auf den Fehlerterm haben. Korreliert beispielsweise die Entfernung zur nächsten Wasch- und Trocknungsanlage zufällig mit einer anderen Variable, die sich im Modell befindet (etwa mit der Temperatur), so verzerrt das den Schätzer für die Modellvariable Temperatur. Korreliert die Entfernung zu Wasch- und Trocknungsanlage mit keiner der anderen unabhängigen Variablen, sind deren Schätzer nicht verzerrt. Die

räumliche Autokorrelation befindet sich dann im Fehlerterm, wodurch die Varianz der Schätzer der unabhängigen Variablen erhöht wird.

Schließlich kann durch eine Fehlspezifikation des Modells künstliche räumliche Autokorrelation auftreten. So kann etwa nicht berücksichtigte Heteroskedastizität der Fehler zu einer zufällig interpretierten, artifiziellen räumlichen Autokorrelation führen.

Liegt räumliche Autokorrelation in einem OLS-Modell vor, so können daher wesentliche Modellannahmen bezüglich der Fehlerterme (Erwartungswert von 0 und Homoskedastizität) (WOOLDRIDGE, 2009, 84ff) verletzt sein, was eine Verzerrung der geschätzten Modellparameter zur Folge haben kann (PLANT 2012, 423ff). Selbiges gilt auch für ein Tobit-Modell auf Basis einer Maximum-Likelihood-Schätzung, da ebenfalls dessen 2 zentrale Modellannahmen (homoskedastische und normalverteilte Fehler) dadurch verletzt werden können (LONG 1997, 206). Um festzustellen, ob ein OLS-Modell oder ein Tobit-Modell nach ML-Schätzung zulässig ist, muss überprüft werden, ob räumliche Autokorrelation vorliegt.

Um räumliche Autokorrelation in einem Regressionsmodell festzustellen und in weiterer Folge zu berücksichtigen, sind eine Nachbarschaftsmatrix ( $W$ ) und eine Spatial-Lag-Variable notwendig (LESAGE und PACE 2009, 8ff): Eine Nachbarschaftsmatrix hat die Dimensionen  $n \times n$  ( $n$  = Anzahl der Beobachtungen) und stellt den räumlichen Bezug der Beobachtungen zueinander her. In der Nachbarschaftsmatrix ist definiert, welche Beobachtungen Nachbarn sind (1) und welche nicht (0). Der räumliche Bezug der Daten zueinander kann auf unterschiedlichste Arten zum Ausdruck gebracht werden.

In ähnlichen Anwendungsfällen wie der vorliegenden Arbeit basieren die Nachbarschaftsdefinitionen auf direkter Nachbarschaft (Gemeinden mit gemeinsamer Grenze) oder einer Schwellenentfernung (Gemeinden innerhalb eines gewissen Entfernungsradius) (siehe z. B. SCHOLZ et al. 2013, ODGAARD et al. 2011 oder LIPPERT et al. 2012). Direkte Nachbarschaft ist eine einfache und intuitive Nachbarschaftsdefinition. Es ergibt sich aber das Problem, dass 2 Beobachtungen trotz räumlicher Nähe keine Nachbarn sind, sofern sie keine gemeinsame Grenze haben. Mit einer distanzbasierten Nachbarschaftsdefinition lässt sich dieses Problem lösen, es muss aber eine geeignete Schwellenentfernung festgelegt werden. Häufig werden die Nachbarschaftsbeziehungen außerdem gewichtet (z.B. auf Basis eines Entfernungskehrwertes), um einen mit Entfernung abnehmenden Effekt simulieren zu können (siehe z. B. BIVAND und BRUNSTAD 2006 ,283). Die am besten geeignete Nachbarschaftsdefinition bzw. -gewichtung hängt stets vom Anwendungsfall ab. Sie sollte die tatsächlichen Bedingungen möglichst gut beschreiben.

Um zu erreichen, dass räumlich nahe gelegene Gemeinden ohne gemeinsame Grenze ebenfalls als Nachbarn gelten, wird in der vorliegenden Arbeit eine distanzbasierte



Nachbarschaftsmatrix gewählt. Es wird außerdem von einem mit Entfernung abnehmenden räumlichen Effekt ausgegangen. Daher wird die Nachbarschaftsmatrix auf Basis eines Entfernungskehrwertes gewichtet.

Bei der Auswahl des Entfernungsschwellenwertes müssen verschiedene Aspekte berücksichtigt werden: In Niederösterreich werden Ölkürbiskerne in Einzelfällen bis zu 30 km zur nächsten Wasch- und Trocknungsanlage transportiert, im Regelfall ist diese Entfernung jedoch geringer (BRANDSTETTER 2013, s.p.). Für Landwirte bzw. Landwirtinnen in der Steiermark ist auch die Nähe zur nächsten Ölmühle von Bedeutung. Ein weiterer Aspekt ist der mögliche Erfahrungs- und Wissensaustausch zwischen benachbarten Landwirten bzw. Landwirtinnen aus unterschiedlichen Gemeinden. Aus methodischer Sicht muss noch berücksichtigt werden, dass jede Gemeinde mindestens eine Nachbargemeinde benötigt (Inselbeobachtungen sind unzulässig).

Die Entfernungen zwischen den Gemeinden werden ermittelt, indem die Luftliniendistanzen von Zentroidpunkten der einzelnen Gemeinden zueinander berechnet werden. Aus praktischer Sicht müssen noch Topographie, Straßenverlauf und die Tatsache, dass es sich um repräsentative Gemeindemittelpunkte handelt, berücksichtigt werden. Die tatsächlich zurückzulegende Entfernung zwischen z.B. einer Wasch- und Trocknungsanlage in einer Gemeinde A und einem landwirtschaftlichen Betrieb in einer Gemeinde B ist daher im Regelfall höher als die Luftlinienentfernung. Die Luftliniendistanz, bei der jede Gemeinde mindestens noch eine Nachbargemeinde aufweist, beträgt 10,8 km. Bei dieser Entfernung hat jede Gemeinde im Mittel rund 14 Nachbargemeinden und die zurückzulegende Fahrstrecke könnte etwa im Bereich von 15-20 km liegen. Bei höheren Luftliniendistanzen steigt die mittlere Anzahl der benachbarten Gemeinden stark an (ca. 26 Nachbarn bei 15 km und 44 Nachbarn bei 20 km)<sup>1</sup>. Unter Berücksichtigung der oben angeführten Aspekte wird daher die Luftliniendistanz von 10,8 km als Entfernungsschwellenwert für die Beschreibung der räumlichen Effekte gewählt.

Standardisiert man eine Nachbarschaftsmatrix  $W$ , indem man die Zeilensummen gleich 1 setzt und multipliziert sie mit einem Vektor von  $n \times 1$  Werten (z. B. mit den Werten der abhängigen Variable  $Y$ ), so erhält man als Ergebnis für jede Variable die Durchschnittswerte der Nachbarn (Spatial-Lags) (LESAGE und PACE 2009, 10):

---

<sup>1</sup> Eine zusammenfassende Darstellung unterschiedlicher getesteter Nachbarschaftsmatrizen findet sich im Anhang.

$$Wy = \begin{bmatrix} 0 & 0,5 & 0 & 0,5 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} y_A \\ y_B \\ y_C \\ y_D \end{bmatrix} = \begin{bmatrix} (y_B + y_D) \div 2 \\ (y_A + y_C) \div 2 \\ (y_B) \div 1 \\ (y_A) \div 1 \end{bmatrix}$$

Räumliche Regressionsmodelle erweitern herkömmliche Regressionsmodelle um eine weitere unabhängige Variable in Form einer solchen Spatial-Lag-Variable, deren Regressionsparameter die Stärke der räumlichen Autokorrelation beschreibt.

Nach der Bildung einer Nachbarschaftsmatrix können Tests auf räumliche Autokorrelation durchgeführt werden. Im Rahmen dieser Arbeit werden 2 Teststatistiken verwendet, um die Residuen des OLS- bzw. des Tobit-Modells auf räumliche Autokorrelation zu testen:

- Moran's I-Test
- Lagrange-Multiplier-Test

Moran's I-Test beruht auf der Nullhypothese keine räumliche Autokorrelation, die gegen die Alternativhypothese räumliche Autokorrelation getestet wird (PLANT 2012, 104ff). Ist seine Teststatistik (Moran's I) signifikant, so kann auf dem gewählten Signifikanzniveau (in der vorliegenden Arbeit dem 5%-Niveau) nicht ausgeschlossen werden, dass räumliche Autokorrelation vorliegt. Der Vorteil des Moran's I-Tests ist seine universelle Anwendbarkeit, sowohl auf das OLS-Modell, als auch auf das Tobit-Modell. Mit Moran's I-Test kann jedoch nicht festgestellt werden, welche Form räumlicher Autokorrelation vorliegt. Diese Frage kann mit dem Lagrange-Multiplier-Test beantwortet werden (ANSELIN 2007, 88 bzw. PLANT 2012, 434f): Bei diesem Test wird ebenfalls von der Nullhypothese „keine räumliche Autokorrelation“ ausgegangen, die gegen die Alternativhypothese „räumliche Autokorrelation“ getestet wird. Dabei werden 4 Teststatistiken berechnet:

- LMlag
- LMerr
- RLMlag
- RLMerr

Ist das Testergebnis von LMlag signifikant, dann liegt räumliche Autokorrelation in der abhängigen Variable vor. Die Annahme dahinter ist, dass die Interaktion von benachbarten wirtschaftenden Akteuren (z. B. Landwirten bzw. Landwirtinnen) zu ähnlichem Verhalten führt. Auf die Kulturpflanzenwahl im Ackerbau übersetzt bedeutet das etwa, dass durch den Austausch von Erfahrungswerten und Wissen, die gemeinsame Nutzung von Ressourcen oder bereits alleine durch die gegenseitige Beobachtung, räumlich nahe gelegene landwirtschaftliche Betriebe sich in deren Kulturpflanzenwahl gegenseitig beeinflussen

können. So bewegen Innovationspioniere, die wenig verbreitete Ackerkulturen erfolgreich kultivieren, andere Betriebe dazu, ebenfalls mit dem Anbau derselben Kultur zu beginnen. Solche räumlichen Wirkungen werden auch als „Nachbarschaftseffekte“ oder „räumliche Spillover-Effekte“ bezeichnet (BICHLER 2006, 42, ANSELIN 2002, 248f). Für diesen Fall eignet sich ein Spatial-Autoregressive -Modell (SAR-Modell) am besten für die Berücksichtigung der räumlichen Autokorrelation. In einem SAR-Modell wird der Regressionsgleichung eine Spatial-Lag-Variable der abhängigen Variable (z.B. durchschnittlicher Ölkürbisanteil der benachbarten Gemeinden) als zusätzliche unabhängige Variable hinzugefügt (LE SAGE und PACE 2009, 32):

$$Y = \rho WY + \alpha + \beta X + U$$

mit

*Y = Vektor der abhängigen Variable mit n beobachteten Werten*

*$\rho$  = Parameter „Rho“, der die Stärke der räumlichen Autokorrelation von Y beschreibt*

*W = Nachbarschaftsmatrix ( $n \times n$ )*

*$\alpha$  = Konstanter Term*

*X = Matrix ( $n \times (k+1)$ ) mit den Ausprägungen der unabhängigen Variablen*

*$\beta$  = Vektor der k Regressionskoeffizienten*

*U = Vektor der n Fehler*

Ist das Testergebnis von LMerr signifikant, so liegt räumliche Autokorrelation im Fehlerterm vor. Dies kann durch räumlich autokorrelierende, jedoch nicht erfassbare oder nicht erfasste unabhängige Variablen hervorgerufen werden (LESAGE und PACE 2009, 32). Mit einem Spatial-Error-Modell (SEM) lässt sich diese Form der räumlichen Autokorrelation beschreiben. Der Fehlerterm (U) wird dabei in einen räumlich autokorrelierenden und einen räumlich nicht autokorrelierenden Teil zerlegt (LESAGE und PACE 2009, 32). Dadurch können zwar keine räumlichen Spillover-Effekte erklärt werden, jedoch lassen sich so die, aufgrund der räumlichen Autokorrelation der Fehler, eventuell ebenfalls verzerrten Regressionsparameter ( $\beta$ ) bereinigen (LESAGE und PACE 2009, 32):

$$Y = \alpha + \beta X + \eta$$

wobei für  $\eta$  gilt

$$\eta = \lambda W\eta + U$$

mit

*$\eta$  = Vektor der n Fehler (räumlich autokorrelierender Anteil)*

*$\lambda$  = Parameter „Lambda“, der die Stärke der räumlichen Autokorrelation von  $\eta$  beschreibt*

*U = Vektor der n Fehler (räumlich nicht autokorrelierender Anteil)*

Es ist auch eine Kombination aus SAR-Modell und SEM-Modell möglich, die von LESAGE und PACE (2009, 32f) als SAC-Modell bezeichnet wird. Im Rahmen dieser Arbeit wird aber auf ein solches Modell nicht näher eingegangen.

Sind die beiden Teststatistiken LMlag und LMerr signifikant, so werden die robusten Teststatistiken (RLMlag und RLMerr) für die Auswahl des Modelles herangezogen (ANSELIN 2003, 12). Ist dann etwa RLMlag signifikant oder signifikanter als RLMerr, so wird ein SAR-Modell gewählt (ANSELIN 2003, 12). Der Lagrange-Multiplier-Test kann nur für das OLS-Modell durchgeführt werden, weshalb für das Tobit-Modell lediglich Moran's I-Test berechnet wird. Es wird davon ausgegangen, dass alle 3 möglichen Arten räumlicher Autokorrelation (Interaktion, Abhängigkeit, Fehlspezifikation) im Modell vorliegen können. Daher ist sowohl räumliche Autokorrelation der abhängigen Variable als auch räumliche Autokorrelation des Fehlerterms denkbar. Um eine eventuelle Fehlspezifikation zu entdecken, wird der RESET-Test angewandt und um die Modelle auf Heteroskedastizität zu testen, wird der Breusch-Pagan Test durchgeführt.

Als Schätzverfahren für die räumlichen Modelle wird für die Erweiterung des OLS-Modells die Maximum-Likelihood-Methode angewandt (siehe LESAGE und PACE 2009, 45 ff). Für die Erweiterung des Tobit-Modells wird ein, auf bayesianischer Statistik beruhendes Schätzverfahren (Markov-Chain-Monte-Carlo-Verfahren, kurz MCMC), angewandt (Details dazu siehe LESAGE und PACE 2009, 123ff). Ein Vorteil des MCMC-Verfahrens ist, dass es robust gegenüber Heteroskedastizität ist (DONFOUET et al. 2013, 18).

### **3.3.4 Modellgleichung und Interpretation der Regressionskoeffizienten**

Neben der bereits in Abschnitt 3.2 beschriebenen Log-Transformation der Variablen Ölkürbisanteil, Körnermaisanteil, Ackerfläche und LF pro Betrieb, werden die Bestimmungsfaktoren für die Verwendung in den Regressionsmodellen noch zentriert. Dazu werden die Mittelwerte von allen  $x_i$  subtrahiert, wodurch die Variablen einen Mittelwert von 0 aufweisen (PLANT 2012, 308). Dies hat 2 Gründe (PLANT 2012, 308):

Der konstante Term  $\alpha$  im Regressionsmodell beschreibt normalerweise den Effekt auf die abhängige Variable, wenn alle unabhängigen Variablen den Wert 0 annehmen. Der Wert 0 ist bei zentrierten Variablen der Mittelwert. Daher gibt  $\alpha$  bei einem Modell mit zentrierten Variablen den Effekt auf die abhängige Variable wieder, wenn alle unabhängigen Variablen ihren Mittelwert annehmen. Die Aussagekraft des konstanten Terms wird dadurch verbessert, da sein Effekt einen realistischeren Wertebereich (Mittelwerte) beschreibt.

Der zweite Grund hängt mit den quadrierten Variablen zusammen: Sind eine lineare und eine quadrierte Variable in einem Modell vorhanden (z.B. Temperatur und Temperatur<sup>2</sup>), so

sind die Werte ihrer geschätzten Regressionsparameter höchst instabil (Parameterinstabilität). Die Ursache dafür ist, dass die beiden Variablen sehr stark miteinander korrelieren (Multikollinearität). Zentriert man die Variablen, ist keine Multikollinearität mehr vorhanden und die geschätzten Regressionsparameter bleiben stabil.

Um die Regressionskoeffizienten in einer geeigneten Größeneinheit interpretieren zu können, werden als weitere Maßnahme einige Variablen skaliert. Die Variable Niederschlag wird durch 10 dividiert und die Variablen GVE pro Ha LF sowie Temperatur mit dem Faktor 10 multipliziert. Die Variablen messen so eine Wertänderung von 1 in einem Intervall von 10 l/m<sup>2</sup> bzw. 0,1 GVE/ha und 0,1°C. Die sich aus diesen Anpassungsmaßnahmen ergebende grundlegende Regressionsgleichung ist folglich am Beispiel des OLS-Modells dargestellt:

$$\begin{aligned} \ln(1 + OelkAnt) = & \alpha + \beta_1 \times ggA_{N\ddot{o}} + \beta_2 \times ggA_{StmkBgld} + \beta_3 \times N\ddot{a}heSpInfrstr \\ & + \beta_4 \times (Temp_{zentr} \times 10) + \beta_5 \times (Temp_{zentr} \times 10)^2 + \beta_6 \times \left(\frac{Ns_{zentr}}{10}\right) \\ & + \beta_7 \times (GVEproHa_{zentr} \times 10) + \beta_8 \times \ln(1 + KmAnt)_{zentr} \\ & + \beta_9 \times [\ln(1 + KmAnt)_{zentr}]^2 + \beta_{10} \times \ln(1 + Ackerfl)_{zentr} \\ & + \beta_{11} \times \ln(1 + LFproBetr)_{zentr} \end{aligned}$$

Die weiteren Regressionsmodelle basieren alle auf dieser Modellgleichung. Im Tobit- sowie im räumlichen Tobit-Modell wird die abhängige Variable  $\ln(1 + OelkAnt)$  durch eine latente Variable ( $Y^*$ ) ersetzt. Für die räumlichen Regressionsmodelle wird der Modellgleichung, je nach Ergebnis der Tests auf räumliche Autokorrelation, noch  $\rho WY$  oder  $\lambda W\eta$  hinzugefügt.

Als marginaler Effekt in einem Regressionsmodell wird der Effekt bezeichnet, den die Wertänderung einer abhängigen Variable um eine Einheit auf die unabhängige Variable hat, während alle anderen unabhängigen Variablen konstant gehalten werden (WOOLDRIDGE 2009, 697). In einem OLS-Modell entspricht der marginale Effekt von  $x_k$  auf  $\hat{y}$  dem geschätzten Regressionskoeffizienten von  $x_k$  ( $\widehat{\beta}_k$ ) (WOOLDRIDGE 2009, 697).

Die Regressionskoeffizienten  $\beta_1 - \beta_7$  des OLS-Modells werden aufgrund der Log-Transformation der Ölkürbisanteile als Semi-Elastizitäten interpretiert (WOOLDRIDGE 2009, 46):

$$\% \Delta \hat{y} = 100 \widehat{\beta}_k \Delta x_i$$

Beträgt der Regressionskoeffizient der Variable  $(GVEproHa_{zentr} \times 10)$  z. B. -0,04 so sinkt der geschätzte Ölkürbisanteil, bei einer Zunahme der GVE/ha von 0,1 GVE/ha, um 4%. Bei

einem ursprünglichen Ölkürbisanteil von 9% wäre das ein Rückgang um 0,36 Prozentpunkte auf 8,64% ( $9 - 9/100 \cdot 4 = 8,64$ ).

Die Regressionskoeffizienten der Log-transformierten Bestimmungsfaktoren werden im OLS-Modell als Elastizitäten interpretiert (WOOLDRIDGE 2009, 46):

$$\% \Delta \hat{y} = \widehat{\beta}_k \% \Delta x_i$$

Beträgt der Regressionskoeffizienten der Variable  $\ln(1 + Ackerfl)_{zentr}$  z. B. 0,3, so bewirkt eine 1%ige Zunahme der Ackerfläche eine Erhöhung des Ölkürbisanteils um 0,3% (z. B. von 15% auf 15,045%).

Die marginalen Effekte der Variablen mit einem linearen ( $\beta_1 x$ ) und einem quadrierten Term ( $\beta_2 x^2$ ) setzen sich aus  $\widehat{\beta}_1 + 2 \widehat{\beta}_2 x$  zusammen (WOOLDRIDGE 2009, 192). Das bedeutet, der Gesamteffekt ist abhängig von dem Wert für  $x$ , der in die obige Formel eingesetzt wird. Häufig wird dafür der Mittelwert verwendet. Da in der vorliegenden Arbeit alle unabhängigen Variablen zentriert sind, beträgt ihr Mittelwert Null. Setzt man den Mittelwert in die Formel ein, verbleibt als marginaler Effekt  $\widehat{\beta}_1$ .

Im Tobit-Modell wird ein nicht-linearer Zusammenhang den Variablen  $X$  und  $Y$  unterstellt, wodurch sich auch die Interpretation der Regressionskoeffizienten verändert (WOOLDRIDGE 2009, 591): Der Term  $\widehat{\beta}_k$  beschreibt zwar noch immer den Einfluss von  $x_{i,k}$  auf die Variable  $\hat{y}_i$ , der Effekt variiert aber mit den Ausprägungen der  $x$ -Werte. Um die Schätzer einer ML-Tobit-Regression und einer OLS-Regression vergleichen zu können, werden erstere mit einem Skalierungsfaktor multipliziert, der zwischen 0 und 1 liegt (WOOLDRIDGE 2009, 591):

$$\frac{1}{n} \sum_{i=1}^n \Phi(x_i \hat{\beta} / \hat{\sigma})$$

mit

$n$  = Anzahl der Beobachtungen

$\Phi$  = kumulative Normalverteilungsfunktion

$x_i$  = unabhängige Variable mit  $n$  Ausprägungen

$\hat{\beta}$  = Schätzer des Regressionskoeffizienten von  $x_i$

$\hat{\sigma}$  = Schätzer der Standardabweichung der Fehler

Multipliziert man  $\widehat{\beta}_k$  mit dem Skalierungsfaktor ergibt das den Average-Partial-Effect (APE). Der APE beschreibt den mittleren Effekt von  $X$  auf  $Y$  (WOOLDRIDGE 2009, 591). Beträgt der Skalierungsfaktor 0,5 und der Regressionskoeffizient der Variable ( $GVEproHa_{zentr} \times 10$ )

-0,04 würde der ursprünglich geschätzte Ölkürbisanteil bei einer Steigerung der GVE/ha um 0,1 im Mittel von z. B. 10% auf 9,8% sinken ( $10 - (100 \times -0,04 \times 0,5) = 9,8$ ).

Die marginalen Effekte der Dummy-Variablen im Tobit-Modell werden nach einer anderen Formel berechnet (BARTUS 2005, 311):

$$APE_k = \frac{1}{n} \sum_{i=1}^n [F(\beta x^i | x_k^i = 1) - F(\beta x^i | x_k^i = 0)]$$

Wendet man diese Formel z. B. auf die Variable *Nähe zu spezieller Infrastruktur* an, so werden zuerst gefittete Werte berechnet, bei denen alle Gemeinden die Ausprägung „Gemeinde liegt in der Nähe von spezieller Infrastruktur“ annehmen. Danach werden nochmals gefittete Werte berechnet, jedoch wird die Dummy-Variable für alle Gemeinden auf die Ausprägung „Gemeinde liegt nicht in der Nähe von spezieller Infrastruktur“ gesetzt. Die Ergebnisse werden voneinander subtrahiert, danach aufsummiert und der Mittelwert daraus berechnet. So erhält man einen mittleren Effekt der Dummy-Variablen, ähnlich dem APE für kontinuierliche Variablen WOOLDRIDGE (2009, 592).

In einem SAR-Modell beeinflusst die Änderung der Variable ( $GVE_{proHa_{zentr}} \times 10$ ) in einer Gemeinde, sowohl den geschätzten Ölkürbisanteil der jeweiligen Gemeinde ( $\hat{y}_i$ ), als auch den Ölkürbisanteil anderer Gemeinden ( $\hat{y}_j$ ). Durch Rückkopplungseffekte, kann sich das wiederum auf den Ölkürbisanteil der ursprünglichen Gemeinde auswirken. Die Stärke dieser Effekte ist von den Ausprägungen aller x-Werte abhängig und wird durch die Matrix  $S_r(W)$  mit den Dimensionen  $n \times n$  beschrieben (LESAGE und PACE 2009, 34ff). LESAGE und PACE (2009, 39) unterscheiden 3 Maßzahlen, die diese Effekte zusammenfassen:

Der „Average Direct Impact“ (ADI) ist der durchschnittliche Einfluss der Änderung der i-ten Beobachtung einer Einflussgröße  $x_r$  ( $x_{ir}$ ) auf die Zielgröße  $\hat{y}_i$ . Der ADI entspricht dem Mittelwert der Diagonalsumme von  $S_r(W)$ .

Der „Average Total Impact“ (ATI) ist der Mittelwert der Zeilen- oder Spaltensumme von  $S_r(W)$  (beide liefern dasselbe Ergebnis). Die Interpretation unterscheidet sich aber in beiden Fällen. Der Mittelwert der Zeilensumme beschreibt den durchschnittlichen Einfluss der Änderung der Einflussgröße  $x_r$  über alle Beobachtungen (z. B. Zunahme der Temperatur in allen Gemeinden um  $0,1^\circ\text{C}$ ) auf eine Zielgröße  $\hat{y}_i$  („ATI to an Observation“). Der Mittelwert der Spaltensumme beschreibt den durchschnittlichen Einfluss der Änderung der Einflussgröße  $x_{ir}$  einer Beobachtung (z. B. Zunahme der Temperatur in einer Gemeinde um  $0,1^\circ\text{C}$ ) auf alle Zielgrößen  $\hat{Y}$  („ATI from an Observation“).

Subtrahiert man den ADI vom ATI, erhält man den „Average Indirect Impact“ (AII). Er beschreibt alle indirekten Einflüsse. So kann etwa durch den AII der Effekt festgestellt

werden, den eine Gemeinde auf alle anderen Gemeinden ausübt, wenn sie die Ausprägung „Gemeinde liegt in der Nähe von spezieller Infrastruktur“ annimmt. Beträgt der All z.B. 1,5, so nimmt der Ölkürbisanteil in allen anderen Gemeinden um insgesamt 150% zu ( $1,5 \cdot 100$ ). Die 150% verteilen sich dabei auf alle anderen Gemeinden. Wie sich die 150% genau auf die einzelnen Gemeinden aufteilen, ist von der gewählten Nachbarschaftsdefinition, der Gewichtung der Nachbarschaftsmatrix und der Stärke des Parameters Rho abhängig (LE SAGE und PAGE 2009, 72f).

Die Regressionskoeffizienten eines SAR-Tobit-Modells werden ähnlich wie im SAR-Modell interpretiert. Die berechneten Effekte (ADI, ATI und All) werden aber zusätzlich mit einem Anpassungsfaktor analog zum Tobit-Modell multipliziert.<sup>2</sup>

Im SEM- bzw. SEM-Tobit-Modell befindet sich der räumliche Regressionsparameter Lambda im Fehlerterm. Daher werden die Regressionskoeffizienten dieser Modelle wie in einem OLS- bzw. Tobit-Modell interpretiert (LE SAGE und PAGE 2009, 41).

Als Statistiksoftware für die Berechnung der Regressionsmodelle wird R (R CORE TEAM 2013, s.p.) in Kombination mit der integrierten Entwicklungsumgebung R-Studio (R-STUDIO 2013, s.p.) verwendet. Folgende R-Pakete werden für die Berechnungen benötigt:

- „aer“ (KLEIBER und ZEILEIS 2008), für Tests für das Tobit-Modell
- „censReg“, (HENNINGSEN 2013), für das Tobit-Modell
- „lmtest“ (ZEILEIS und HOTHORN 2002), für den RESET-Test und den Breusch-Pagan-Test
- „MASS“ (RIPLEY et al. 2012), für die Variablenauswahl auf Basis des AIC
- „spatialprobit“ (WILHELM und GODINHO DE MATOS 2013), für die räumliche Erweiterung des Tobit-Modells
- „spdep“ (BIVAND 2013), für die Bildung von Nachbarschaftsmatrizen und die räumliche Erweiterung des linearen Modells
- „vgam“ (YEE 2013), für das Tobit-Modell

---

<sup>2</sup> Für die Berechnungen von ADI, All und ATI im SAR-Tobit-Modell wurde aufgrund fehlender Implementierung eine eigene Funktion in R geschrieben, mit Formeln aus WOOLDRIDGE 2009, 591 und LESAGE und PACE (2009, 294 sowie 302ff). Das dazugehörige R-Skript findet sich im Anhang.



## 4 Ergebnisse

In Abschnitt 4.1 werden die Ergebnisse der 4 berechneten Modelle vorgestellt und miteinander verglichen. Zur Beurteilung der Ergebnisse der Regressionsmodelle sowie der statistischen Tests wird in dieser Arbeit ein Signifikanzniveau von 5% verwendet, das im wirtschafts- und sozialwissenschaftlichen Bereich üblicherweise angewandt wird (WOOLDRIDGE 2009, 123). Auf die Berechnung und Interpretation der marginalen Effekte wird in Abschnitt 4.2 eingegangen. Zum Abschluss wird in Abschnitt 4.3 eine räumliche Analyse der geschätzten Ölkürbisanteile und deren Abweichungen von den tatsächlichen Ölkürbisanteilen durchgeführt.

### 4.1 Regressionskoeffizienten, Gütemaße und statistische Tests

Die Ergebnisse der 4 berechneten Regressionsmodelle sind in Tabelle 2 dargestellt. Im oberen Teil der Tabelle befinden sich die Schätzer der Regressionskoeffizienten. Die Regressionskoeffizienten beschreiben nicht direkt die marginalen Effekte der unabhängigen Variablen auf den Ölkürbisanteil. Aus den Vorzeichen kann man jedoch erkennen, ob der jeweilige Bestimmungsfaktor positiv (+) oder negativ (-) auf den Ölkürbisanteil wirkt.

Zur Beurteilung der Modellgüte finden sich unterhalb der Regressionskoeffizienten noch das  $R^2$  und das AIC. Das  $R^2$  im Tobit-, SAR-, und SAR-Tobit-Modell ist der quadrierte Korrelationskoeffizient zwischen den beobachteten und den geschätzten Werten der abhängigen Variable<sup>3</sup>. Die Interpretation dieses alternativen  $R^2$  bleibt annähernd gleich: Je näher es bei 1 liegt, desto höher ist die Korrelation zwischen gefitteten und beobachteten Werten von  $Y$  und demzufolge auch der Modell-Fit.

Im unteren Teil der Tabelle sind die berechneten Teststatistiken der Modelle dargestellt. Die Signifikanzniveaus der Regressionskoeffizienten und der Teststatistiken sind entsprechend der Kategorisierung am Fuß der Tabelle gekennzeichnet.

---

<sup>3</sup> Details siehe WOOLDRIDGE 2009, 592.

**Tabelle 2: Ergebnisse der Regressionsmodelle: Regressionskoeffizienten, Gütemaße und Teststatistiken**

| <b>Abhängige Variable: Log-transformierte Ölkürbisanteile</b> |                 |                 |                 |                        |
|---|-----------------|-----------------|-----------------|------------------------|
| <b>Anzahl der Beobachtungen: 1347 Gemeinden</b>               |                 |                 |                 |                        |
| <b>Unabhängige Variablen</b>                                  | <b>OLS</b>      | <b>Tobit</b>    | <b>SAR</b>      | <b>SAR-Tobit</b>       |
| Konstante   | 0,0893**        | -0,7868***      | 0,0241          | -0,3169***             |
| g.g.A.-Gebiet Nö.   | 0,8578***       | 1,4690***       | 0,2504***       | 0,5470***              |
| g.g.A.-Gebiet Stmk.-Bgld.                                     | 1,3199***       | 2,1258***       | 0,2891***       | 0,7505***              |
| Nähe zu spezieller Infrastruktur                              | 0,5496***       | 0,4799***       | 0,2011***       | 0,1865*                |
| Temperatur <sub>zentr</sub> × 10                              | 0,0158***       | 0,0234***       | 0,0034          | 0,0051                 |
| (Temperatur <sub>zentr</sub> × 10) <sup>2</sup>               | 0,0007*         | 0,0015**        | 0,0003          | 0,0008                 |
| Niederschlag <sub>zentr</sub> / 10                            | 0,0077***       | 0,0104*         | 0,0051**        | 0,0086*                |
| GVE pro ha LF <sub>zentr</sub> × 10                           | -0,0239***      | -0,0542***      | -0,0194***      | -0,0425***             |
| ln (1 + Körnermaisanteil) <sub>zentr</sub>                    | 0,1677***       | 0,2824***       | 0,0838***       | 0,1649***              |
| [ln (1 + Körnermaisanteil) <sub>zentr</sub> ] <sup>2</sup>    | 0,0529***       | 0,0430          | 0,0130          | -0,0113                |
| ln (1 + Ackerfläche) <sub>zentr</sub>                         | 0,0749***       | 0,2262***       | 0,0576***       | 0,1933***              |
| ln (1 + LF pro Betrieb) <sub>zentr</sub>                      | -0,2405***      | -0,3659***      | -0,1139**       | -0,2177**              |
| Sigma   | 0,5712          | 0,8848***       | 0,4451          | 0,6603***              |
| Rho   | / <sup>a)</sup> | / <sup>a)</sup> | 0,7230***       | 0,6667***              |
| <b>Gütemaße</b>   |                 |                 |                 |                        |
| R <sup>2</sup> <sup>b)</sup>                                  | 0,71            | 0,73            | 0,82            | 0,82                   |
| AIC   | 2.327,71        | 2.290,10        | 1.782,94        | 1.952,26               |
| <b>Teststatistiken</b>  |                 |                 |                 |                        |
| RESET-TEST  | 95,54***        | 6,34*           | 205,63***       | 18,48*** <sup>c)</sup> |
| Breusch-Pagan-Test  | 275,70***       | 96,24***        | 196,38***       | 73,67*** <sup>c)</sup> |
| RLMlag (Inv. Distanz 10,8 km)                                 | 107,39***       | / <sup>a)</sup> | / <sup>a)</sup> | / <sup>a)</sup>        |
| RLMerr (Inv. Distanz 10,8 km)                                 | 29,58***        | / <sup>a)</sup> | / <sup>a)</sup> | / <sup>a)</sup>        |
| Moran's I Residuen (Inv. Distanz 10,8 km)                     | 0,35***         | 0,54***         | 0,01            | 0,02                   |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

Signifikanzniveaus: \*\*\*: 0 - 0,001 (0,1%-Niveau); \*\*: > 0,001 – 0,01 (1%-Niveau); \*: > 0,01 – 0,05 (5%-Niveau)

a) Für Felder, die mit „/“ gekennzeichnet sind, konnte kein Wert berechnet werden.

b) Das R<sup>2</sup> wurde für das OLS-Modell nach der herkömmlichen Formel und für die anderen Modelle als quadriertes Korrelationskoeffizient von  $\hat{Y}$  und Y berechnet (Details siehe WOOLDRIDGE 2009, 592)

c) Der Reset- und Breusch-Pagan-Test des SAR-Tobit-Modells beruhen auf einer ML-Schätzung eines SAR-Modells mit der latenten Variable Y\* des SAR-Tobit-Modells als abhängige Variable.

Im OLS-Modell sind alle unabhängigen Variablen signifikant. Die Vorzeichen der Regressionskoeffizienten entsprechen außerdem, mit Ausnahme der Variablen  $(Temperatur_{zentr} \times 10)^2$  und  $[\ln(1 + Körnermaisanteil)_{zentr}]^2$ , den getroffenen Annahmen. Das berechnete  $R^2$  in Höhe von 0,71 besagt, dass das OLS-Modell die Daten gut beschreibt. Die Teststatistiken zeigen aber, dass im OLS-Modell einige Modellannahmen verletzt werden: Sowohl die Ergebnisse des RESET- (95,54) als auch des Breusch-Pagan-Tests (275,70) sind höchst signifikant, was klar darauf hindeutet, dass sowohl Fehlspezifikation als auch Heteroskedastizität im OLS-Modell vorliegen. Auch die Tests auf räumliche Autokorrelation liefern eindeutig signifikante Ergebnisse. Fehlspezifikation und Heteroskedastizität sind vermutlich zumindest teilweise auf die zahlreichen Nullen in der abhängigen Variable (Log-transformierte Ölkürbisanteile) zurückführbar. Daher wird ein Tobit-Modell berechnet.

Die Werte der Regressionskoeffizienten des Tobit-Modells sind, mit Ausnahme der Variablen *Nähe zu spezieller Infrastruktur* und  $[\ln(1 + Körnermaisanteil)_{zentr}]^2$ , größer als im OLS-Modell. Das ist ein erstes Indiz dafür, dass der wahre Modellzusammenhang mit dem OLS-Modell aufgrund der zahlreichen Nullen nicht korrekt geschätzt wird. Die Vorzeichen der Koeffizienten entsprechen aber noch immer jenen des OLS-Modells. Mit Ausnahme von  $[\ln(1 + Körnermaisanteil)_{zentr}]^2$  sind alle Regressionskoeffizienten noch mindestens auf dem 5%-Niveau signifikant. Das  $R^2$  (0,73) und das AIC (2.290,10) deuten darauf hin, dass der Fit des Tobit-Modells im Vergleich zum OLS-Modell etwas besser ist.

Das Ergebnis des RESET-Tests (6,34) ist nur mehr knapp auf dem 5%-Niveau signifikant. Das deutet darauf hin, dass der funktionale Zusammenhang von  $X$  und  $Y$  mit der latenten Variable  $Y^*$  im Tobit-Modell weitaus besser beschrieben wird. Das Ergebnis des Breusch-Pagan-Tests weist noch immer auf Heteroskedastizität hin, auch wenn die Teststatistik von 275,7 auf 96,24 gesunken ist. Der Wert von 96,24 ist aber noch klar auf dem 0,1%-Niveau signifikant. Außerdem zeigt das ebenfalls signifikante Ergebnis von Moran's I-Test (0,54), dass auch noch räumliche Autokorrelation im Tobit-Modell vorliegt.

Aufgrund der signifikanten Ergebnisse des RESET- und Breusch-Pagan-Tests kann allerdings nicht eindeutig ausgeschlossen werden, dass die vorhandene räumliche Autokorrelation eventuell auf einer Fehlspezifikation des Modells beruht<sup>4</sup>. Da aber keine zusätzlichen Informationen für weitere unabhängige Variablen zur Verfügung stehen, die diesen Umstand aufklären könnten, werden auf Basis der Ergebnisse des Lagrange-Multiplier-Tests (OLS-Modell) und des Moran's I-Tests (OLS- und Tobit-Modell) räumliche Regressionsmodelle gebildet.

---

<sup>4</sup> Siehe die Anmerkungen dazu in Abschnitt 3.3.3.

Sowohl die Teststatistik  $RLMlag$  (107,39), als auch die Teststatistik  $RLMerr$  (29,58) des Lagrange-Multiplier-Tests sind auf dem 0,1%-Niveau signifikant. Da der Wert von  $RLMlag$  höher ist, werden als räumliche Erweiterungen Modelle mit einer Spatial-Lag-Variable von  $Y$  (SAR- bzw. SAR-Tobit-Modell) gebildet.

Das SAR-Modell wird mit einer ML-Schätzung berechnet. Mit diesem Modell wird räumliche Autokorrelation berücksichtigt, die zahlreichen Nullen im Datensatz verzerren aber noch immer die Schätzergebnisse. Dadurch werden wieder die Modellannahmen in Bezug auf die Fehler (Normalverteilung und Homoskedastizität) verletzt. Die Ergebnisse dieses Modells sind daher, analog zu denen des OLS-Modells, mit Vorbehalt zu behandeln.

Die Werte der Regressionskoeffizienten im SAR-Modell sind durchwegs kleiner als im OLS- bzw. Tobit-Modell. Das liegt daran, dass ein größerer Teil der Varianz der Ölkürbisanteile durch die Ölkürbisanteile benachbarter Gemeinden erklärt wird. Der Parameter  $Rho$  beträgt 0,72 und ist auf dem 0,1%-Niveau signifikant. Die geschätzten Ölkürbisanteile benachbarter Gemeinden hängen daher im SAR-Modell stark positiv zusammen.

Der Modell-Fit ist zwar besser als im OLS- und Tobit-Modell, aber die Teststatistiken des RESET-Tests (205,63) und des Breusch-Pagan-Test (196,38) sind eindeutig auf dem 0,1%-Niveau signifikant und liegen deutlich über denen des Tobit-Modells. Moran's  $I$  beträgt 0,01 und ist nicht mehr auf dem 5%-Niveau signifikant. Es ist also keine räumliche Autokorrelation mehr in den Residuen des SAR-Modells vorhanden.

Mit dem SAR-Tobit-Modell sollen die positiven Eigenschaften des Tobit-Modells (bessere Modellierung des funktionalen Zusammenhangs von  $X$  und  $Y$  und geringere Heteroskedastizität) mit denen des SAR-Modells (Berücksichtigung räumlicher Autokorrelation) vereint werden. Die Ergebnisse der letzten Spalten in Tabelle 2 zeigen, dass die Vorzeichen der Regressionskoeffizienten, mit Ausnahme der Variable ( $Temperatur_{zentr} \times 10$ )<sup>2</sup> noch immer den Annahmen entsprechen. Die Werte der Schätzer liegen zwischen denen des OLS- und Tobit-Modells. Diese Veränderung beruht wiederum auf dem hoch signifikanten und positiven Parameter  $Rho$ , der einen Teil des ursprünglichen Erklärungsgehaltes der anderen Modellvariablen aufnimmt. Der räumliche Zusammenhang ist etwas schwächer als im SAR-Modell, aber nach wie vor sehr stark ( $Rho \sim 0,67$ ).

Die 3 Dummy-Variablen sind auch in diesem Modell signifikant und lassen einen positiven Einfluss auf den geschätzten Ölkürbisanteil erkennen. Die Variable *Nähe zu spezieller Infrastruktur* ist aber nur noch auf dem 5%-Niveau signifikant. Die beiden Temperatur-Variablen sind nicht mehr auf dem 5%-Niveau signifikant, ebenso wie die quadrierten Körnermaisanteile. Die Variablen  $Niederschlag_{zentr} / 10$  und  $\ln(1 + LF \text{ pro Betrieb})_{zentr}$  sind nur noch auf dem 5%- bzw. 1%-Niveau signifikant. Die restlichen Variablen weisen weiterhin

ein Signifikanzniveau von 0,1% auf.  $R^2$  (0,82) und AIC (1.952.26) deuten auf einen ähnlich guten Fit wie beim SAR-Modell hin.

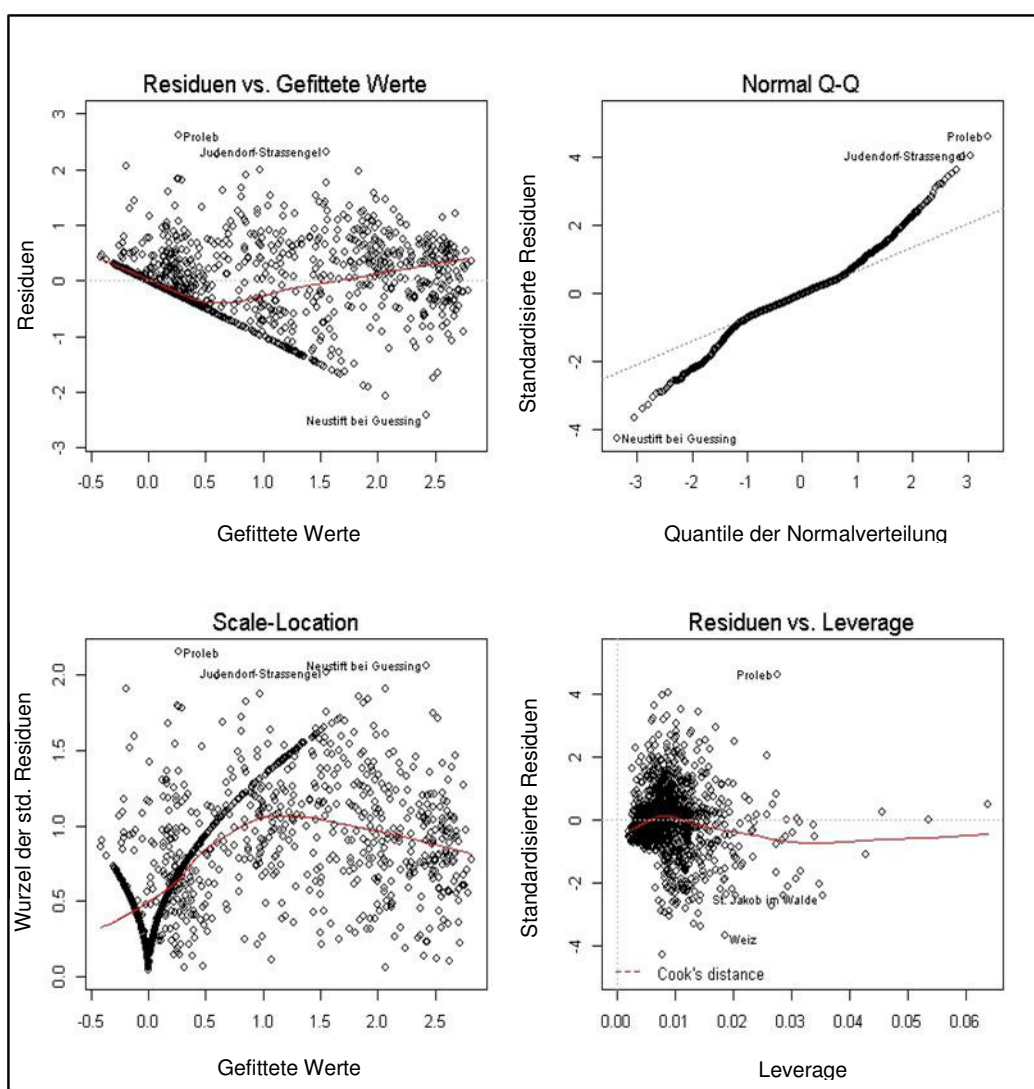
Der RESET- und Breusch-Pagan-Test konnten aufgrund mangelnder Implementierung in R nicht für das SAR-Tobit-Modell berechnet werden. Um zumindest ungefähre Richtwerte für die beiden Testgrößen zu erhalten, wird zunächst ein SAR-Modell mit der latenten Variable  $Y^*$  des SAR-Tobit-Modells als abhängige Variable berechnet (SAR- $Y^*$ -Modell). Aus theoretischer Sicht wird so derselbe Modellzusammenhang wie im SAR-Tobit-Modell geschätzt. Der RESET- sowie der Breusch-Pagan-Test werden dann auf Basis des SAR- $Y^*$ -Modells berechnet. Diese Vorgehensweise wird mit den nahezu identen Ergebnissen dieses Modells im Vergleich zum SAR-Tobit-Modell gerechtfertigt. Die Ergebnisse der beiden Modelle weichen aufgrund der unterschiedlichen angewandten Schätzverfahren (ML und MCMC) leicht voneinander ab.

Das Ergebnis des RESET-Tests ist auf dem 0,1%-Niveau signifikant. Der Wert von 18,48 ist im Vergleich zum Tobit-Modell (6,34) zwar höher, aber immer noch deutlich geringer als im OLS- bzw. SAR-Modell. Die Teststatistik des Breusch-Pagan-Tests ist im Vergleich zum Tobit Modell von rund 96 auf ca. 73 gesunken. Das Ergebnis ist aber immer noch eindeutig auf dem 0,1%-Niveau signifikant. Moran's I beträgt 0,02. Das bedeutet es ist auf dem 5%-Niveau keine räumliche Autokorrelation mehr in den Residuen des SAR-Tobit-Modells vorhanden. Die Ergebnisse zeigen, dass das SAR-Tobit-Modell, im Vergleich zu den anderen Modellen, die Daten am besten beschreibt. Die Ergebnisse des RESET- und Breusch-Pagan-Tests deuten jedoch noch immer auf leichte Fehlspezifikation und starke Heteroskedastizität hin.

Um diese beiden Probleme näher zu untersuchen, wird eine grafische Analyse der Ergebnisse mit Hilfe von Diagnostikplots durchgeführt. In Abbildung 9 sind 4 Diagnostikplots für das OLS-Modell dargestellt. Die beiden linken Grafiken, in denen die gefitteten Werte gegen die Residuen bzw. die standardisierten Residuen abgebildet sind, sollten idealerweise eine zufallsverteilte Punktwolke und einen horizontalen Verlauf der roten Linie zeigen. Stattdessen sind eindeutige Muster erkennbar, die durch die Gemeinden entstehen, in denen kein Ölkürbis angebaut wird. Das Muster nimmt in der linken oberen Grafik eine Trichterform (Hinweis auf Heteroskedastizität) und in der linken unteren Grafik eine V-Form (Hinweis auf Fehlspezifikation) an.

In der rechten oberen Grafik sind die Quantile der Residuen des Regressionsmodells gegen die Quantile einer Normalverteilung abgebildet. Liegen die Residuen auf der gepunkteten Linie, so folgen sie einer Normalverteilung. Die starke Abweichung von der Linie an den beiden Enden der Quantile, deutet jedoch darauf hin, dass die Residuen keiner Normalverteilung folgen.

Die rechte untere Grafik stellt den Effekt etwaiger Ausreißer auf die Regressionsgleichung dar. Gemeinden mit ungewöhnlichen X-Werten (z.B. überdurchschnittlich große Summe der Ackerflächen) und gleichzeitig hohen Residuen, können die Werte der geschätzten Regressionskoeffizienten stark verändern. Als Maßzahl für solche Beobachtungen kann die Cook's-Distance herangezogen werden (CHATERJEE und HADI 1988, 117). Eine Gemeinde gilt in der Grafik dann als problematisch, wenn sie eine Cook's-Distance erreicht, die größer als 0,5 ist (= hoher Leverage und hohes Residuum). Diese Grenze wird in R durch ein rot gestricheltes Band gekennzeichnet. Das rote Band ist in der Grafik nicht sichtbar, da keine Beobachtung in der Nähe der Begrenzung liegt. Das bedeutet, es gibt keine problematischen Ausreißer in den Daten.

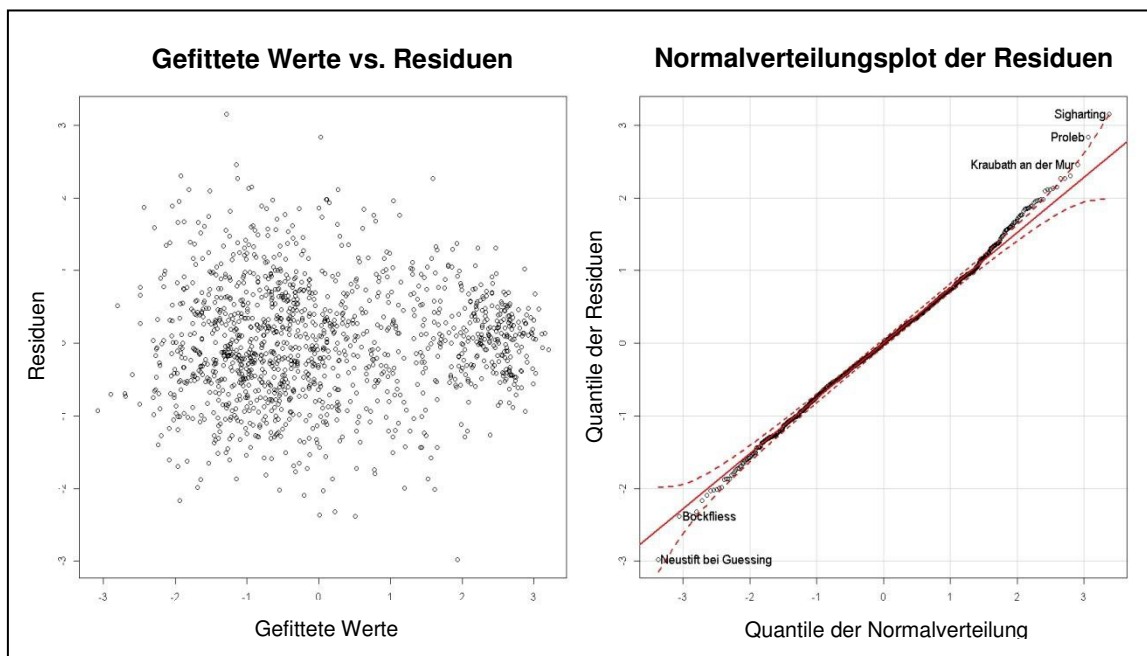


**Abbildung 9: Diagnostikplots OLS-Modell**

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

Als Vergleich zum OLS-Modell, sind in Abbildung 10 ein Plot der gefitteten Werte ( $\widehat{Y}^*$ ) gegen die Residuen ( $Y^* - \widehat{Y}^*$ ) und ein Normalverteilungsplot der Residuen des SAR-Tobit-Modells dargestellt. Dadurch, dass die latente Variable  $Y^*$ , im Gegensatz zur beobachteten Variable  $Y$ , negative Werte anstatt der Nullen annimmt, lässt sich kein klares Muster mehr in der Punktwolke erkennen.

Die Residuen folgen außerdem approximativ einer Normalverteilung. Nur im rechten oberen Bereich des Normalverteilungsplots liegen einige Gemeinden knapp außerhalb des 95%igen Konfidenzintervalls, das als rot gestricheltes Band gekennzeichnet ist. Das betrifft unter anderem die Gemeinden Sigharting (Oberösterreich), Proleb (Steiermark, außerhalb des g.g.A.-Gebietes) und Kraubath an der Mur (Steiermark, außerhalb des g.g.A.-Gebietes). In diesen Gemeinden wird der Ölkürbisanteil vom Modell relativ stark unterschätzt. Am linken unteren Rand des Normalverteilungsplots der Residuen befindet sich die Gemeinde Neustift bei Güssing (g.g.A.-Gebiet im Burgenland). In dieser Gemeinde wird der Ölkürbisanteil durch das Modell überschätzt. Die Abweichung liegt allerdings noch knapp innerhalb des 95%igen Konfidenzintervalls.



**Abbildung 10: Diagnostikplots SAR-Tobit-Modell**

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

## 4.2 Marginale Effekte

Da aus den bisherigen Ergebnissen hervorgeht, dass das Tobit- sowie das SAR-Tobit-Modell die Modellannahmen am besten erfüllen, werden nachfolgend nur die marginalen Effekte

dieser beiden Modelle untersucht<sup>5</sup>. Eine Übersicht der marginalen Effekte des OLS- und SAR-Modells findet sich im Anhang.

In Tabelle 3 befinden sich die Average Partial Effects (APEs) des Tobit-Modells. Die Regressionskoeffizienten des Tobit-Modells werden dafür mit dem berechneten APE-Skalierungsfaktor ( $\sim 0,5$ ) multipliziert. Um die Semi-Elastizitäten direkt interpretieren zu können, werden die APEs der nicht logarithmierten unabhängigen Variablen außerdem mit dem Faktor 100 multipliziert.

Die APEs der Dummy-Variablen beschreiben die prozentuale Differenz des Ölkürbisanteils im Vergleich zur Basiskategorie (kein g.g.A.-Gebiet und keine Nähe zu spezieller Infrastruktur). So ist der geschätzte Ölkürbisanteil in Gemeinden im niederösterreichischen g.g.A.-Gebiet im Mittel um 94,12% höher als in Gemeinden außerhalb des g.g.A.-Gebietes. Im g.g.A.-Gebiet in der Steiermark und dem Burgenland ist dieser Effekt noch stärker – der Ölkürbisanteil ist dort im Mittel 145,68% höher. Diese Ergebnisse zeigen bereits, dass die Zweiteilung des g.g.A.-Gebietes sinnvoll zu sein scheint, da sich die beiden Effekte in ihrem Ausmaß eindeutig unterscheiden. In Gemeinden, die sich in der Nähe von spezieller Infrastruktur befinden, ist der Ölkürbisanteil 26,49% höher als in den anderen Gemeinden. Die Annahme, dass diese 3 Effekte maßgeblich die Höhe des Ölkürbisanteils bestimmen, wird demgemäß bestätigt. Durch die anderen Modellvariablen sind die Effekte außerdem hinsichtlich möglicher klimatischer (Temperatur, Niederschlag, Körnermaisanteil) sowie größenstrukturbedingter (Ackerfläche, LF pro Betrieb) Einflüsse bereinigt.

Die anderen Bestimmungsfaktoren wirken im Vergleich dazu weniger stark auf den Ölkürbisanteil. Die beiden positiven APEs der Temperatur-Variablen (1,17 und 0,07) bedeuten, dass ein positiver Temperatureffekt modelliert wird, der mit steigender Temperatur zunimmt. Erhöht sich der Mittelwert der Temperatur im Untersuchungsgebiet etwa um  $0,1^{\circ}\text{C}$ , so steigt der geschätzte Ölkürbisanteil dadurch um  $1,17 + 2 \times 0,07 \times 0 = 1,17\%$ . Der Effekt von  $10 \text{ l/m}^2$  mehr Niederschlag auf den geschätzten Ölkürbisanteil ist mit  $0,52\%$  noch geringer. Die GVE pro ha LF haben einen negativen Effekt auf den Ölkürbisanteil. Steigen die GVE pro ha LF um 0,1, sinkt der geschätzte Ölkürbisanteils um  $2,7\%$ .

---

<sup>5</sup> Für weitere Details zur Berechnung und Interpretation von APE, ADI, All und ATI siehe Kapitel 3.3.4.



Tabelle 3: Marginale Effekte Tobit-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$<br>N = 1347 Gemeinden |                              |             |
|---|------------------------------|-------------|
| Unabhängige Variablen   | Average Partial Effect (APE) | Signifikanz |
| g.g.A.-Gebiet Nö.   | 94,12                        | ***         |
| g.g.A.-Gebiet Stmk.-Bgld.   | 145,68                       | ***         |
| Nähe zu spezieller Infrastruktur  | 26,49                        | ***         |
| Temperatur <sub>zentr</sub> $\times 10$                                     | 1,17                         | ***         |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup>                     | 0,07                         | **          |
| Niederschlag <sub>zentr</sub> / 10  | 0,52                         | *           |
| GVE pro ha LF <sub>zentr</sub> $\times 10$                                  | -2,70                        | ***         |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$                           | 0,14                         | ***         |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$                       | 0,02                         |             |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$                                | 0,11                         | ***         |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$                             | -0,18                        | ***         |

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\*: > 0,001 – 0,01 (1%-Niveau); \*: > 0,01 – 0,05 (5%-Niveau)

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

Steigt der mittlere Körnermaisanteil um ein Prozent, so erhöht sich der Ölkürbisanteil dadurch im Mittel um  $0,14 + 2 \times 0,02 \times \ln(0 + 1) = 0,14\%$ . Der marginale Effekt der Variable  $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$  beträgt 0,11. Sinkt die Ackerfläche um 1%, nimmt der geschätzte Ölkürbisanteil um 0,11% ab. Das Vorzeichen dieses APEs ist zwar positiv, da die Ackerflächen aber tendenziell rückläufig sind, wird er umgekehrt beschrieben. In erster Linie dient diese Variable jedoch als Kontrollvariable, um den Einfluss anderer Variablen bei konstant gehaltener Ackerfläche beschreiben zu können. Die durchschnittliche LF pro Betrieb hat einen negativen Effekt auf den Ölkürbisanteil. Steigt sie um 1%, sinkt der geschätzte Ölkürbisanteils um 0,18%.

Die marginalen Effekte des SAR-Tobit-Modells sind in Tabelle 4 dargestellt. Sie werden in direkte (ADI), indirekte (AII) und gesamte (ATI) mittlere Effekte unterteilt. Die Effekte werden ebenfalls mit einem Skalierungsfaktor multipliziert, um zu berücksichtigen, dass in einem Teil der Gemeinden kein Ölkürbis angebaut wird. Der Skalierungsfaktor ist im Vergleich zum Tobit-Modell etwas geringer und beträgt ca. 0,46. Die nicht Log-transformierten Variablen werden außerdem mit dem Faktor 100 multipliziert. Ein Problem besteht bei der Berechnung der marginalen Effekte für die Dummy-Variablen. Es ist keine Formel für die Berechnung von ADI, AII und ATI für diskrete Variablen in einem SAR-Tobit-Modell verfügbar. Die in Tabelle 4

dargestellten Effekte werden nach der Formel für kontinuierliche Variablen berechnet. Dadurch werden die marginalen Effekte der Dummy-Variablen tendenziell unterschätzt.

Der ADI misst den mittleren Effekt von Änderungen eines Bestimmungsfaktors in einer Gemeinde, auf den Ölkürbisanteil in derselben Gemeinde. Daher lässt er sich von seiner Wirkungsweise mit dem APE des Tobit-Modells vergleichen. Der Vergleich zeigt, dass der ADI tendenziell geringer ist als der APE. Die Berücksichtigung räumlicher Effekte scheint den Bestimmungsfaktoren im SAR-Tobit-Modell einen Teil ihres ursprünglichen Erklärungsgehaltes entzogen zu haben.

Der All ist im Vergleich zum ADI höher. Es gilt aber zu berücksichtigen, dass der All die Summe der Einflüsse auf alle anderen Gemeinden im Untersuchungsgebiet ist. Hat eine Gemeinde etwa 3 Nachbargemeinden, so verteilt sich z.B. der All von 16,38% für *Nähe zu spezieller Infrastruktur*, gemäß der Gewichtung (1/Distanz), größtenteils auf die 3 benachbarten Gemeinden. Ein kleinerer Teil der 16,38% wird von diesen Gemeinden wiederum an ihre Nachbarn weitergegeben.

Wird in einer Gemeinde *i* spezielle Infrastruktur errichtet, so steigt der geschätzte Ölkürbisanteil in allen Gemeinden um rund 25% (ATI from an Observation). Der ATI ist in diesem Fall als die Summe der Einflüsse auf alle Gemeinden (inklusive der Gemeinde *i*) zu verstehen. Der Vergleich zwischen ATI und APE lässt erkennen, dass die mittleren Gesamteffekte im SAR-Tobit-Modell tendenziell größer als im Tobit-Modell sind. Dadurch, dass räumliche Effekte modelliert werden, verändern sich die Ergebnisse und deren Interpretation maßgeblich. Es werden neben direkten Effekten auch indirekte Agglomerationseffekte berücksichtigt.

Eine genauere Interpretation von ADI, All und ATI der 3 Dummy-Variablen erscheint nicht sinnvoll, da die korrekten Werte nicht berechnet werden können. Die Lage im g.g.A.-Gebiet sowie die Nähe zu spezieller Infrastruktur scheinen aber statistisch signifikante, stark positive Effekte sowohl auf den Ölkürbisanteil der betreffenden Gemeinden, als auch auf den Ölkürbisanteil benachbarter Gemeinden zu haben. Die Effekte der anderen Variablen sind im Vergleich dazu deutlich geringer.

Tabelle 4: Marginale Effekte SAR-Tobit-Modell (ADI, AII und ATI)

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$<br>N = 1347 Gemeinden |       |       |        |             |
|---|-------|-------|--------|-------------|
| Unabhängige Variablen   | ADI   | AII   | ATI    | Signifikanz |
| g.g.A.-Gebiet Nö.   | 27,20 | 48,02 | 75,23  | ***         |
| g.g.A.-Gebiet Stmk.-Bgld.   | 37,32 | 65,89 | 103,21 | ***         |
| Nähe zu spezieller Infrastruktur  | 9,28  | 16,38 | 25,66  | *           |
| Temperatur <sub>zentr</sub> $\times 10$                                     | 0,25  | 0,45  | 0,70   |             |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup>                     | 0,04  | 0,07  | 0,11   |             |
| Niederschlag <sub>zentr</sub> / 10  | 0,43  | 0,76  | 1,19   | *           |
| GVE pro ha LF <sub>zentr</sub> $\times 10$                                  | -2,11 | -3,73 | -5,85  | ***         |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$                           | 0,08  | 0,15  | 0,23   | ***         |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$                       | -0,01 | -0,01 | -0,02  |             |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$                                | 0,10  | 0,17  | 0,27   | ***         |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$                             | -0,11 | -0,19 | -0,30  | **          |

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\*: > 0,001 – 0,01 (1%-Niveau); \*: > 0,01 – 0,05 (5%-Niveau)

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013 Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

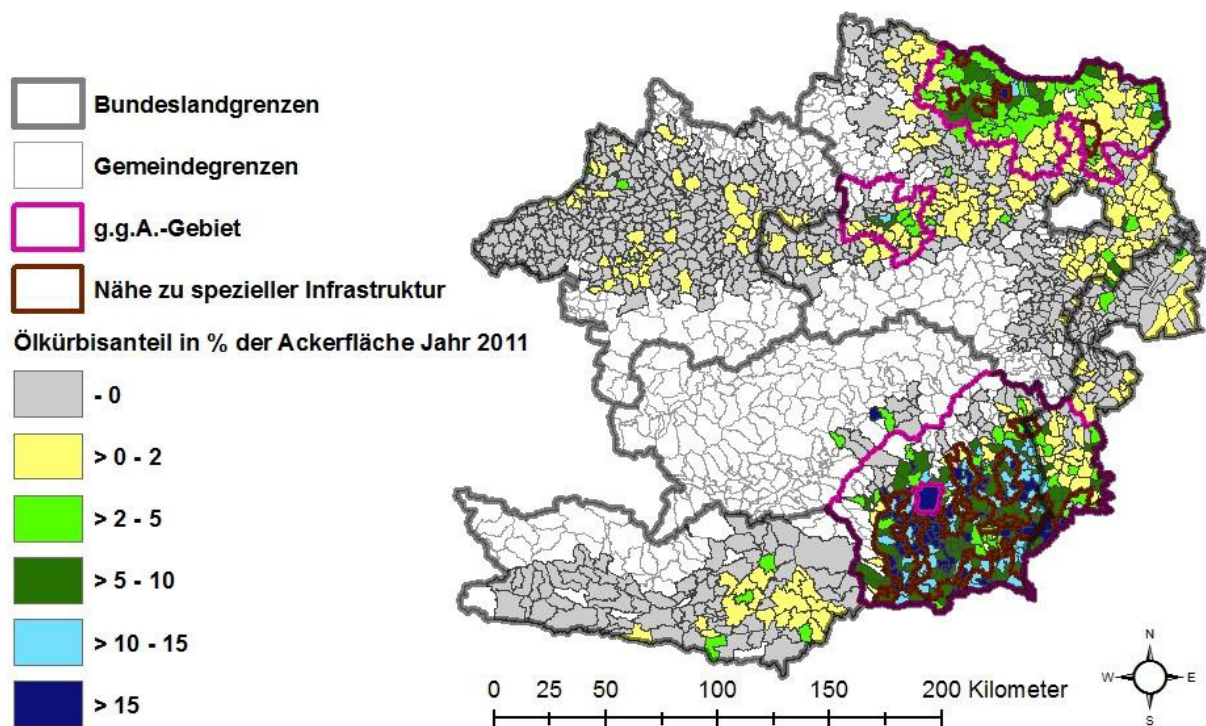
Die Interpretation von AII und ATI der Klimavariablen (Temperatur, Niederschlag sowie Körnermaisanteil) erscheint zunächst etwas abwegig. Es ist davon auszugehen, dass klimatische Veränderungen nicht nur in einer Gemeinde, sondern in allen Gemeinden, bzw. in räumlich zueinander nahe gelegenen Gemeinden auftreten. Daher erscheint für diese Variablen die Interpretation der Ergebnisse nach der „ATI to an Observation“-Perspektive angebracht. Erhöht sich die durchschnittliche Temperatur im Untersuchungsgebiet um  $0,1^\circ\text{C}$ , so würde der geschätzte Ölkürbisanteil einer einzelnen Gemeinde  $i$  um  $0,70 + 2 \times 0,11 \times 0 = 0,7\%$  zunehmen. Steigt der Körnermaisanteil um 1%, so würde sich der geschätzte Ölkürbisanteil um  $0,23 + 2 \times -0,02 \times \ln(0+1) = 0,23\%$  erhöhen. Nimmt der durchschnittliche Niederschlag in allen Gemeinden um  $10\text{l/m}^2$  zu, so würde der geschätzte Ölkürbisanteil der  $i$ -ten Gemeinde um 1,19% steigen.

Die Ackerfläche und die durchschnittliche LF pro Betrieb werden ähnlich interpretiert. Sinkt die Ackerfläche in allen Gemeinden im Schnitt um 1%, so würde sich der geschätzte Ölkürbisanteil in der  $i$ -ten Gemeinde um 0,27% verringern. Eine Zunahme der LF pro Betrieb aller Gemeinden um 1%, hätte eine Reduktion von 0,3% des geschätzten Ölkürbisanteils in einer Gemeinde  $i$  zur Folge.

### 4.3 Vergleich der gefitteten Werte und Residuen

Die bisherigen Ergebnisse zeigen, welche Bestimmungsfaktoren einen statistisch signifikanten Einfluss auf den Ölkürbisanteil haben und wie stark dieser Effekt ist. In diesem Abschnitt wird die räumliche Verteilung der Schätzwerte untersucht. Eine Analyse der gefitteten Werte und ein Vergleich mit den tatsächlichen Ölkürbisanteilen sowie eine Analyse der Abweichungen (Residuen), sollen mögliche Stärken und Schwächen der Modelle aufzeigen. Darüber hinaus können so eventuell Regionen mit vorhandenem Potential für den Ölkürbisbau gefunden werden.

In Abbildung 11 sind nochmals die Ölkürbisanteile des Untersuchungsgebietes für das Jahr 2011 auf Gemeindeebene dargestellt. Die Kategorisierung unterscheidet sich leicht von der in Abbildung 5, um die geschätzten und tatsächlichen Ölkürbisanteile direkt miteinander vergleichen zu können.



**Abbildung 11: Ölkürbisanteile in Prozent der Ackerfläche nach Gemeinden, Jahr 2011**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)

Mit den gefitteten Werten der Regressionsmodelle werden Log-transformierte Ölkürbisanteile nach der Formel  $OelkAntLog = \ln(1 + OelkAnt)$  geschätzt. Bevor mit der Analyse dieser Werte begonnen werden kann, müssen zunächst die gefitteten Werte in Schätzwerte der Ölkürbisanteile umgerechnet werden. Durch Umformung der obigen Formel und Auflösung nach  $OelkAnt$  ergibt sich:  $OelkAnt = \exp(OelkAntLog) - 1$ . WOOLDRIDGE (2009, 211) weist

jedoch darauf hin, dass dadurch die Werte für  $Y$  systematisch unterschätzt werden. Er schlägt zur Beseitigung dieses Problems die Multiplikation der Schätzer mit einem Anpassungsfaktor  $\gamma$  vor, der durch einen Schätzer ( $\hat{\gamma}$ ) approximiert werden kann. Die Formel für die Berechnung von  $\hat{\gamma}$  lautet (WOOLDRIDGE 2009, 212)

$$\hat{\gamma} = \frac{1}{n} \sum_{i=1}^n \exp(\hat{u}_i)$$

mit

$\hat{\gamma}$  = Schätzer des Anpassungsfaktors  $\gamma$

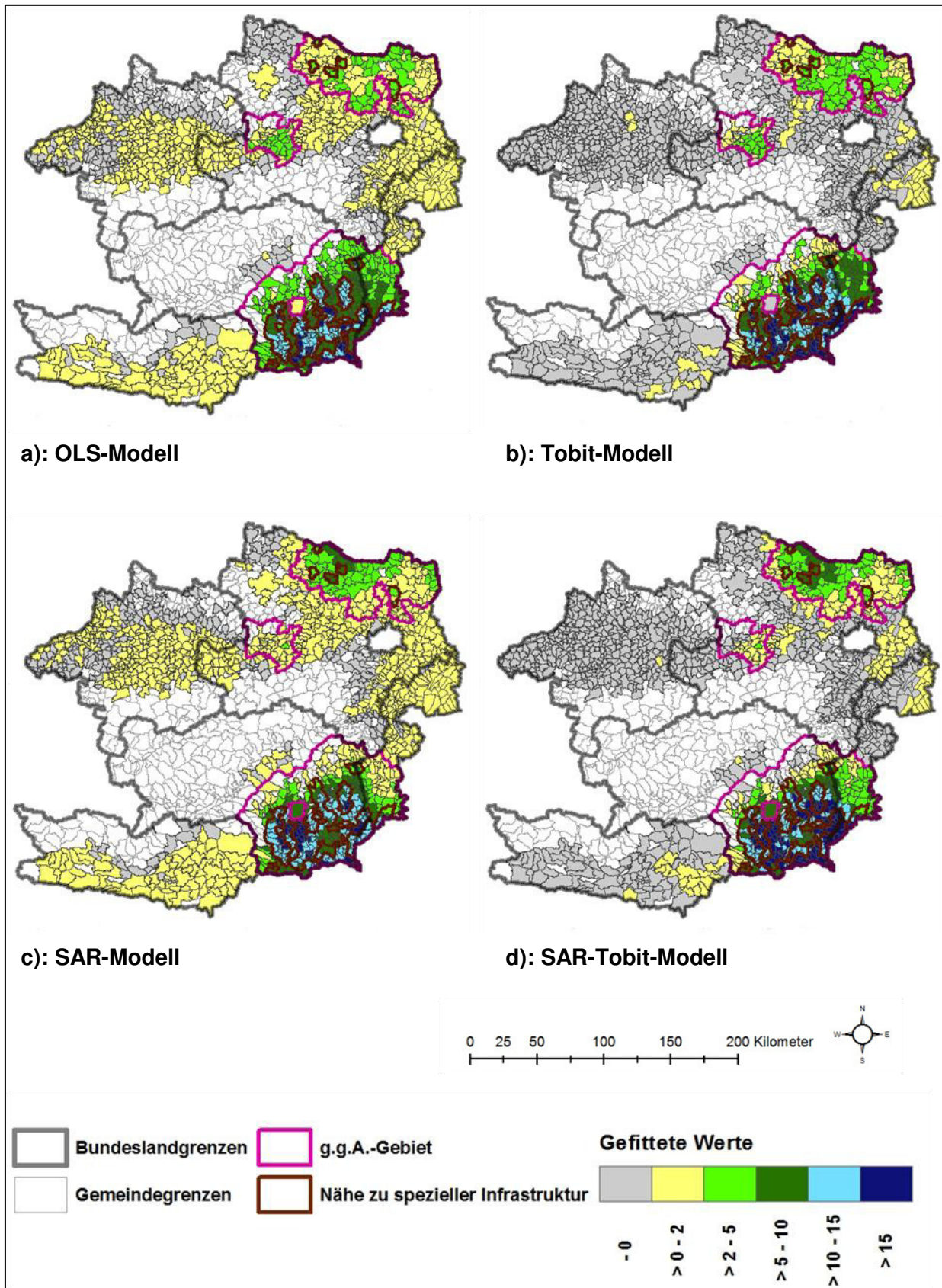
$\hat{u}$  = Residuen des Regressionsmodells

Die nach dieser Formel berechneten Schätzer für die Anpassungsfaktoren betragen  $\sim 1,19$  (OLS-Modell),  $\sim 1,32$  (Tobit Modell),  $\sim 1,15$  (SAR-Modell) und  $\sim 1,24$  (SAR-Tobit-Modell). Zieht man von den Ölkürbisanteilen in Prozent der Ackerfläche die Schätzer der Ölkürbisanteile ab, erhält man die Abweichungen. Die so ermittelten Werte sind in Abbildung 12 und Abbildung 13 dargestellt. Dadurch können die gefitteten Werte mit den tatsächlichen Werten verglichen, und eine räumliche Analyse der Residuen der 4 Regressionsmodelle durchgeführt werden. Um den Effekt des g.g.A.-Gebietes sowie die Nähe zu spezieller Infrastruktur sichtbar zu machen, sind in den Abbildungen die Grenzen des g.g.A.-Gebietes sowie Gemeinden innerhalb eines 5 km-Entfernungsradius zu einer Gemeinde mit spezieller Infrastruktur (Ölmühle oder Wasch- bzw. Trocknungsanlage) gekennzeichnet.

Vergleicht man Abbildung 11 sowie Abbildung 12 a und c ist eindeutig ersichtlich, dass im OLS- sowie im SAR-Modell die Werte von Gemeinden, in denen kein Ölkürbis angebaut wird, tendenziell leicht überschätzt werden. Im Tobit- (b) und SAR-Tobit-Modell (d) wird der funktionale Zusammenhang von  $X$  und  $Y$  mit Hilfe der Variable  $Y^*$  besser beschrieben. Den negativen Schätzwerten von  $Y^*$  wird einfach der Wert 0 zugewiesen. Dadurch ergibt sich, dass in sehr vielen Gemeinden ein Ölkürbisanteil von 0% geschätzt wird.

Eine Gegenüberstellung der nicht-räumlichen und räumlichen Modelle macht deutlich, dass letztere neben den direkten Effekten der Bestimmungsfaktoren auch indirekte Agglomerationseffekte mit in die Berechnungen einbeziehen. Räumlich nahe gelegene Gemeinden, in denen der Ölkürbisanteil höher ist, verbinden sich in diesen Modellen zu Clustern. Gut sichtbar ist auch, dass sich die Cluster in den räumlichen Modellen auch vermehrt über die Grenzen von Gebieten hinaus erstrecken, die laut Definition in der „Nähe von spezieller Infrastruktur“ liegen.

Die Gemeinden mit den höchsten geschätzten Ölkürbisanteilen befinden sich im g.g.A.-Gebiet in der Steiermark und dem Burgenland, vor allem in der Nähe von spezieller Infrastruktur. Die Schätzwerte im niederösterreichischen g.g.A.-Gebiet sind geringer. Die Gemeinden mit den höchsten geschätzten Ölkürbisanteilen, liegen dort in der Gegend um Retz und Pulkau. Außerhalb des g.g.A.-Gebietes wird im Tobit- bzw. im SAR-Tobit-Modell auch ein geringer Ölkürbisanteil für einige Gemeinden im nördlichen Burgenland, dem Marchfeld sowie in Teilen Kärntens und Oberösterreichs geschätzt. Während im Tobit-Modell dabei mehr klimatische (Temperatur, Niederschlag, Körnermaisanteil) bzw. größenstrukturelle (Ackerfläche, LF pro Betrieb) Aspekte eine Rolle zu spielen scheinen, werden im SAR-Tobit-Modell auch die Ölkürbisanteile benachbarter Gemeinden zur Schätzung des Ölkürbisanteils einer Gemeinde herangezogen und somit Agglomerationseffekte berücksichtigt.



**Abbildung 12: Geschätzte Ölkürbisanteile in Prozent der Ackerfläche nach Gemeinden, Jahr 2011**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)

In Abbildung 13 sind die Abweichungen der Modellschätzwerte von den tatsächlichen Ölkürbisanteilen dargestellt. Es bestätigt sich, dass mit dem OLS- bzw. SAR-Modell die Ölkürbisanteile außerhalb des g.g.A.-Gebietes tendenziell leicht überschätzt und innerhalb des g.g.A.-Gebietes eher unterschätzt werden. Mit dem Tobit-Modell werden hingegen vor allem die Ölkürbisanteile im burgenländischen g.g.A.-Gebiet überschätzt. Dies könnte an nicht messbaren hemmenden Faktoren liegen, andererseits können die Abweichungen auch als eventuell vorhandene Potentiale für den Ölkürbisbau gedeutet werden. Im SAR-Tobit-Modell ist diese Tendenz schwächer, da es niedrigere Ölkürbisanteile benachbarter Gemeinden berücksichtigt. Im steirischen g.g.A.-Gebiet sind die Residuen sehr heterogen. Es sind sowohl hohe positive als auch negative Abweichungen zu erkennen. Auch hier kann mit den vorhandenen Daten keine eindeutige Aussage getroffen werden.

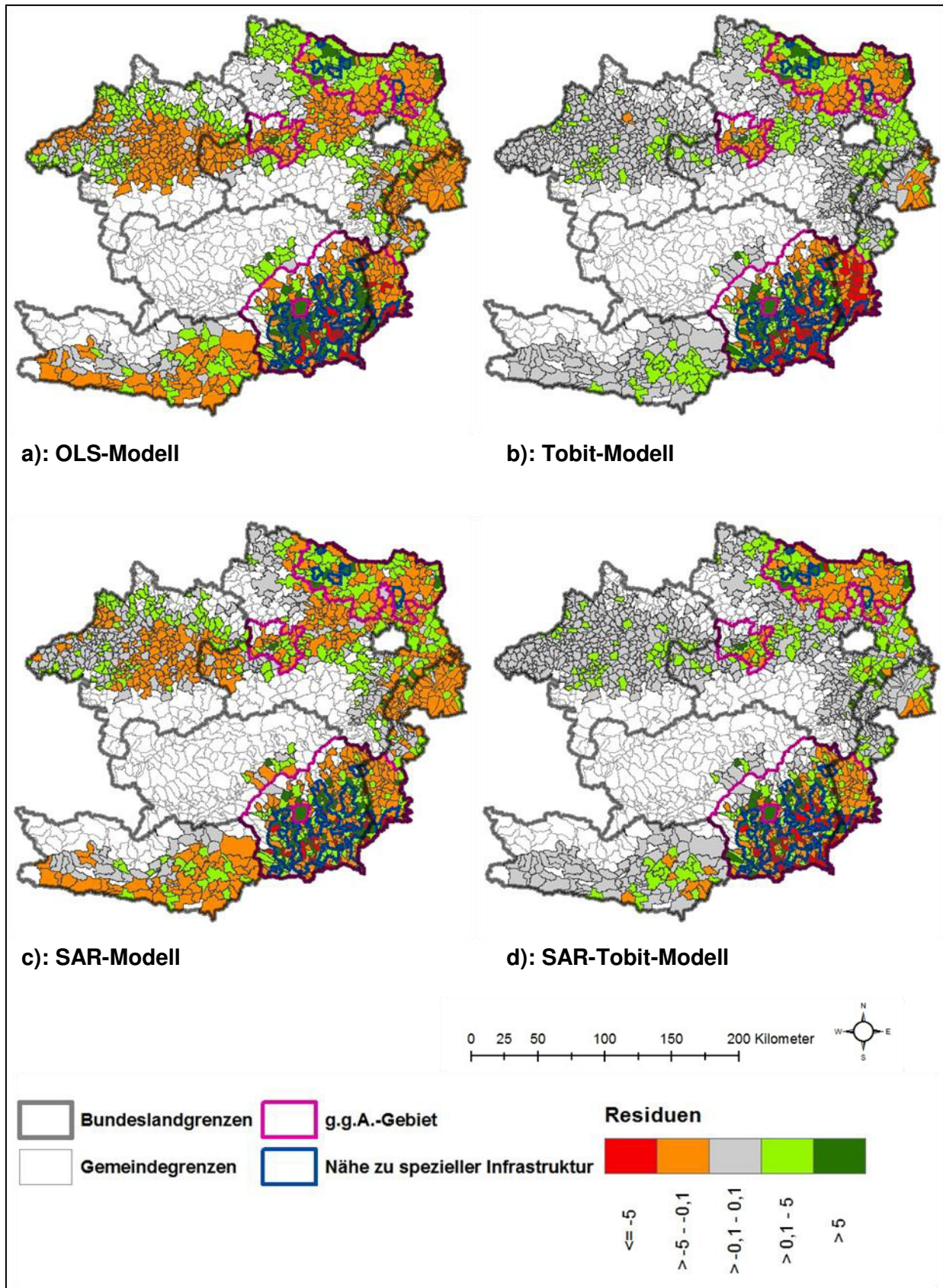
Abschließend sind in Tabelle 5 noch die geschätzten Ölkürbisbauflächen der 1.347 untersuchten Gemeinden für das Jahr 2011 dargestellt. Dafür werden die geschätzten Ölkürbisanteile der 4 Regressionsmodelle mit den Ackerflächen je Gemeinde multipliziert und das Ergebnis aufsummiert. Den negativen Schätzwerten des OLS- bzw. SAR-Modells wird der Wert 0 zugewiesen, was deren geschätzte Ölkürbisflächen noch leicht nach oben korrigiert. Es zeigt sich, dass mit dem OLS- und SAR-Modell zu niedrige Ölkürbisflächen geschätzt werden. Im Gegensatz dazu liegen das Tobit- und SAR-Tobit-Modell relativ nahe an den Anbauflächen des Jahres 2011. Die positiven und negativen Abweichungen heben sich also bei diesen beiden Modellen in Summe ungefähr auf.

**Tabelle 5: Geschätzter Flächenbedarf für den Ölkürbisbau 2011**

| <b>Regressionsmodell</b> | <b>Ölkürbisfläche<br/>2011 in ha</b> | <b>Geschätzte<br/>Ölkürbisfläche<br/>2011 in ha</b> | <b>Differenz in ha</b> |
|--------------------------|--------------------------------------|---|------------------------|
| OLS-Modell               | 25.564,51                            | 23.063,22   | 2.501,29               |
| SAR-Modell               | 25.564,51                            | 23.482,10   | 2.082,41               |
| Tobit-Modell             | 25.564,51                            | 25.724,69   | -160,18                |
| SAR-Tobit-Modell         | 25.564,51                            | 25.822,10   | -257,59                |

*Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)*





**Abbildung 13: Abweichungen der Schätzer von den Ölkürbisanteilen in Prozent der Ackerfläche nach Gemeinden, Jahr 2011**

Quelle: Eigene Darstellung, Daten aus der GEDABA (2012), Standorte der Ölmühlen von der Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013), Standorte von Wasch- und Trocknungsanlagen vom Maschinenring (2010)

## 5 Diskussion und Schlussfolgerungen

Im Rahmen dieser Arbeit soll mit Hilfe (räumlicher) Regressionsmodelle festgestellt werden, welche Bestimmungsfaktoren einen statistisch signifikanten Einfluss auf den Ölkürbisanbau in ausgewählten Gebieten Österreichs haben. Zusätzlich sollen die Wirkungsweise der Bestimmungsfaktoren und die Rolle möglicher räumlicher Effekte geklärt werden. Die empirische Untersuchung erstreckt sich über 1.347 Gemeinden, die mit Querschnittsdaten für das Jahr 2011 analysiert werden. Ausgehend von einem OLS-Modell werden in weiterer Folge ein Tobit-, ein SAR- sowie ein SAR-Tobit Modell berechnet, um schließlich ein Regressionsmodell zu erhalten, das sowohl dem hohen Anteil an Nullwerten in der Zielvariable Rechnung trägt, als auch vorhandene räumliche Autokorrelation in den Daten berücksichtigt. Mit räumlichen Regressionsmodellen können räumliche Effekte wie z. B. Agglomerationseffekte modelliert werden. Räumliche Effekte können im Ölkürbisanbau etwa durch Erfahrungs- und Wissensaustausch, überbetriebliche Maschinenkooperationen und eine gewisse räumliche Bindung zu be- und verarbeitenden Betrieben auftreten.

Aus den Ergebnissen geht hervor, dass in allen berechneten Regressionsmodellen Modellannahmen verletzt werden. Im Rahmen dieser Arbeit kann daher nicht zweifelsfrei festgestellt werden, ob die Modellergebnisse gültig sind und es können folglich nur bedingte Aussagen über die Bestimmungsfaktoren des Ölkürbisanbaus in Österreich getätigt werden. Für die Beurteilung der Ergebnisse wird nachfolgend das SAR-Tobit-Modell gewählt, da es noch am ehesten die Modellannahmen erfüllt. Als Signifikanzniveau wird das 5%-Niveau verwendet.

### 5.1 Methode

Die Untersuchung ist auf die Gemeindeebene beschränkt. Es wurde bereits in Abschnitt 3.1 darauf hingewiesen, dass eine Analyse auf Ebene der Entscheidungsträger für den Ölkürbisanbau (Landwirte und Landwirtinnen) aufgrund fehlender Daten nicht möglich war. Eine aggregierte Betrachtungsebene kann jedoch aufgrund des „Modifiable Area Unit Problems“ sowie des „Ecological Fallacy Problems“ zu ungültigen Rückschlüssen auf die einzelbetriebliche Ebene führen (ANSELIN 2002, 261f). Es kann daher im Rahmen dieser Arbeit nicht überprüft werden, ob die festgestellte positive räumliche Autokorrelation auf Gemeindeebene, auch auf der Ebene landwirtschaftlicher Betriebe vorliegt.

Das Untersuchungsgebiet umfasst 1.347 österreichische Gemeinden, in denen der Ölkürbisanbau aus klimatischer Sicht möglich erscheint. Es stehen jedoch Daten für 1.611 Gemeinden zu Verfügung. Diese Einschränkung des Untersuchungsgebietes kann wesentliche Auswirkungen auf die Ergebnisse haben. Ein größerer Anteil an Nullen in der

abhängigen Variable führt dazu, dass der Konstante Term sinkt und die Schätzer der Regressionskoeffizienten steigen (LONG 1997, 189ff). Durch die Verkleinerung des Untersuchungsgebietes von 1.611 auf 1.347 Gemeinden sinkt der Anteil der Nullen in der abhängigen Variable von ca. 60% auf rund 52%. Die Schätzer der Regressionsparameter sind dadurch vermutlich geringer als in Modellen mit 1.611 Gemeinden. Die Ergebnisse dieser Arbeit sind daher nur für die 1.347 Gemeinden des Untersuchungsgebietes gültig. Der gewählte Ansatz wird trotzdem als zielführend erachtet, da so nur Gemeinden untersucht werden, in denen der Ölkürbisanbau relevant sein könnte.

Die zeitliche Ausdehnung der Analyse beschränkt sich auf das Jahr 2011. Speziell da der Ölkürbisanbau seit Inkrafttreten des Herkunftsschutzes eine dynamische Entwicklung durchlaufen hat und nach wie vor durchläuft, wäre eine Analyse mit Paneldaten über einen längeren Zeitraum wünschenswert (z.B. beginnend ab dem EU-Beitritt Österreichs 1995). So könnten, neben den verwendeten Faktoren, beispielsweise auch Preisentwicklungen der Ölkürbiskerne und konkurrierender Ackerkulturen mit in die Analyse einbezogen werden. Erhöhen sich etwa aufgrund einer positiven Preisentwicklung die Ölkürbisflächen, so wird dieser Effekt in einem Modell mit Querschnittsdaten anderen Faktoren zugerechnet (z.B. dem g.g.A.-Gebiet), da mit diesem Modell eine Preisänderung nicht erfasst werden kann. Die vorliegenden Modellergebnisse dürfen daher nur als statische Eindrücke gesehen werden, die auf einem gegebenen Preisniveau beruhen.

AUTANT-BERNARD und LESAGE (2009) demonstrieren etwa die Anwendbarkeit eines räumlichen Tobit-Modells mit gepoolten Paneldaten. Sie untersuchen räumliche Spillover-Effekte im Wissensbildungsprozess unterschiedlicher französischer Industriesektoren, der die Generierung von Patenten zur Folge hat und stoßen dabei auf positive räumliche Autokorrelation in der abhängigen Variable. Die Analyse erfolgt auf Ebene von 94 Departements und umfasst den Zeitraum von 1992 bis 2000. Im Rahmen der vorliegenden Arbeit ist eine Analyse des Ölkürbisanbaus mit Paneldaten nicht möglich, da keine geeignete Statistik-Software zur Verfügung steht.

Als Bestimmungsfaktoren des Ölkürbisanbaus werden in der vorliegenden Arbeit die Lage einer Gemeinde im g.g.A.-Gebiet in Niederösterreich bzw. in der Steiermark oder dem Burgenland, die Nähe zu spezieller Infrastruktur, die Temperatur, das Quadrat der Temperatur, der Niederschlag, die Anzahl der GVE/ha LF, der Körnermaisanteil, das Quadrat des Körnermaisanteils, die Ackerfläche und die LF pro Betrieb verwendet.

Auf den Ölkürbisanteil, den Körnermaisanteil, das Quadrat des Körnermaisanteils, die Ackerfläche und die LF pro Betrieb wird eine Log-Transformation nach der Formel  $\ln(1+x)$  angewandt, um sie an eine Normalverteilung anzunähern und den Effekt von Ausreißern zu verringern. Wie WOOLDRIDGE (2009, 192) anmerkt, werden dabei durch Addition der

Konstante 1 die Schätzwerte der Regressionsparameter leicht verzerrt, wobei die Verzerrungen umso größer werden, desto mehr Nullen sich im Datensatz befinden. Da rund 52% der Ölkürbisanteile den Wert Null annehmen ist davon auszugehen, dass die Ergebnisse verzerrt sind. Die einzige Alternative wäre gewesen, keine Log-Transformationen durchzuführen, was sich jedoch wiederum negativ auf die Erfüllung der Modellannahmen ausgewirkt hätte. Daher wird dieser Umstand in Kauf genommen.

Ein Aspekt, auf den bereits im Ergebnisteil hingewiesen wurde, sind die Ergebnisse des RESET- und Breusch-Pagan-Tests des SAR-Tobit-Modells. Diese dürfen nur als approximative Richtgrößen gesehen werden, da sie nur indirekt ermittelt werden konnten. Das entsprechende R-Skript, in dem die durchgeführten Berechnungen für die Teststatistiken dokumentiert sind, befindet sich im Anhang.

## 5.2 Ergebnisse

Ein direkter Vergleich der Ergebnisse mit denen anderer Studien ist nicht möglich, da nach dem Wissensstand des Autors noch keine vergleichbare Analyse für den Ölkürbisbau durchgeführt wurde. Es existieren jedoch Untersuchungen, die sich mit den Bestimmungsfaktoren des Maisanbaus befassen. Diese Studien werden hinsichtlich der verwendeten Bestimmungsfaktoren sowie Modellspezifikationen mit den Ergebnissen des SAR-Tobit-Modells verglichen.

ODGAARD et al. (2011) verwenden für die Schätzung Log-transformierter Maisanbauflächen ein SAR-Modell mit Maize-Heating-Units, Niederschlag, Rindviehdichte, Sandanteil im Boden sowie landwirtschaftlich genutzter Fläche als Bestimmungsfaktoren. Die räumliche Ebene der Untersuchung sind 10 x 10 km Rasterdaten für ganz Dänemark, wobei für die Jahre 1999-2008 Modelle mit Querschnittsdaten gebildet werden. Sie stellen positive signifikante Effekte von Temperatur, Rindviehdichte sowie Sandanteil fest.

LIPPERT et al. (2012) ziehen die durchschnittliche Jahrestemperatur, den durchschnittlichen Jahresniederschlag, die Bodenzahl, den Grünlandanteil, die GVE-Dichte, die Biogasanlagen-Dichte und eine Dummy-Variable für Städte zur Erklärung der räumlichen Verteilung des Maisanbaus in Baden-Württemberg heran. Die Untersuchung erfolgt auf Bezirksebene mit Daten für das Jahr 2007, unter Verwendung eines Spatial Error Modells mit inverser Distanzgewichtung der Nachbarschaftsmatrix  $W$ . Als Ergebnisse der Untersuchung verzeichnen sie positive Temperatureffekte sowie einen positiven räumlichen Regressionsparameter der Residuen.

SCHOLZ et al. (2013) untersuchen die Bestimmungsfaktoren des Silomaisanbaus in Bayern. Die Analyse wird auf Landkreisebene für das Jahr 2010 unter Verwendung eines

heteroskedastierobusten SARAR-Modelles (Kombination aus SAR- und SEM-Modell) durchgeführt. Die Nachbarschaftsmatrix bilden sie auf Basis direkter Nachbarschaft. Sie können keinen eindeutigen Nachweis für einen positiven räumlichen Regressionsparameter der abhängigen Variable (Silomaisanteil in % der LF) erbringen, entdecken aber positive räumliche Autokorrelation der Residuen. Als weitere signifikante Bestimmungsfaktoren geben die Autoren unter anderem die Biogasanlagenleistung sowie die Rindviehdichte an.

LANGYINTUO und MEKURIA (2008) untersuchen Nachbarschaftseffekte bei der Adaption neuer Maissorten durch Landwirte und Landwirtinnen in Mosambik. Als methodischen Zugang wählen sie ein SAR-Tobit-Modell auf Basis von Querschnittsdaten, welche die Autoren im Zuge einer Befragung auf einzelbetrieblicher Ebene von 300 Landwirten und Landwirtinnen erheben. Sie stellen unter anderem positive räumliche Autokorrelation in der abhängigen Variable fest.

Im Vergleich zu einigen dieser Studien zeigt das in dieser Arbeit gebildete SAR-Tobit-Modell keinen signifikanten Temperatureffekt. Die beiden Temperatur-Variablen sind zwar positiv, jedoch eindeutig nicht mehr auf dem 5%-Niveau signifikant. Da das Untersuchungsgebiet auf Basis klimatischer Bedingungen eingegrenzt ist, könnte das den Temperatureffekt abgeschwächt haben. Es ist aber auch möglich, dass die Klima-Proxyvariablen für den Körnermaisanteil die klimatische Eignung für den Ölkürbisbau besser beschreiben als die Temperatur.

Die Variable Niederschlag hat im SAR-Tobit-Modell einen positiven Effekt auf den geschätzten Ölkürbisanteil und ist noch auf dem 5%-Niveau signifikant. Die im Durchschnitt höheren Niederschläge im g.g.A.-Gebiet in der Steiermark decken sich mit dieser Annahme. Auch in der Literatur wird auf die günstigen Bedingungen für den Ölkürbisbau im warmen und feuchten illyrischen-Klima der südöstlichen Steiermark und dem südlichen Burgenland hingewiesen (KULMER und WEBER 1990, 14).

Das negative Vorzeichen des Bestimmungsfaktors GVE/ha LF im SAR-Tobit-Modell entspricht den im Rahmen dieser Arbeit getroffenen Annahmen. Die Variable ist außerdem auf dem 0,1%-Niveau signifikant. Es lässt sich jedoch nicht genauer eruieren, ob dieser Effekt primär auf einer arbeitszeitbedingter Konkurrenz zwischen Tierhaltung und Ölkürbisbau oder einer indirekten Beschreibung der Topografie durch die GVE pro ha LF beruht (positive Korrelation der GVE pro ha LF mit dem Grünlandanteil in Höhe von 0,43).

WOOLDRIDGE (2009, 681) empfiehlt für Regressionsanalysen mit aggregierten Beobachtungen, eine Variable für deren „Größe“ zu verwenden. In dieser Arbeit wurde dafür die verfügbare Ackerfläche je Gemeinde gebildet. Mit dieser Variable kann der Ölkürbisanteil an der Summe der Ackerfläche je Gemeinde hinsichtlich etwaiger Größeneffekte kontrolliert

werden. Der auf dem 0,1%-Niveau signifikante positive Effekt der Ackerfläche bedeutet, dass der Ölkürbisanteil in Gemeinden mit mehr Ackerfläche tendenziell größer ist, als in Gemeinden mit geringerer Ackerfläche.

Die durchschnittliche LF pro Betrieb stellt die Größenstruktur der Betriebe dar. Der Effekt der Betriebsgröße ist negativ und auf dem 1%-Niveau signifikant. Dieses Ergebnis deckt sich mit der Annahme, dass in Gemeinden mit einer kleineren Betriebsstruktur mehr Ölkürbis angebaut wird. Die durchschnittliche LF pro Betrieb liegt im gesamten Untersuchungsgebiet bei 24,94 ha, im g.g.A.-Gebiet der Steiermark und dem südlichen Burgenland bei 12,75 ha, aber im niederösterreichischen g.g.A.-Gebiet bei 38,21 ha. Es ist also denkbar, dass vor allem die Direktvermarktung von Kürbiskernöl für viele kleinere Betriebe eine Möglichkeit bietet, eine höhere Wertschöpfung pro ha zu erzielen. Die größeren Betriebe, vor allem in Niederösterreich haben eine andere Strategie: Sie kultivieren Ölkürbis überwiegend im Vertragsanbau. Das zeigt, dass auch in Regionen mit einer größeren Betriebsstruktur mehr Ölkürbis angebaut wird, wenn fördernde Standortfaktoren vorhanden sind (Lage im g.g.A.-Gebiet und Nähe zu spezieller Infrastruktur).

In den oben genannten Studien werden meist natürliche Standortfaktoren (Temperatur, Niederschlag, Bodenzahl) zur Erklärung des Maisanbaus verwendet. Neben diesen Punkten ist der Ölkürbisanteil einer Gemeinde vor allem davon abhängig, ob sich diese im g.g.A.-Gebiet und eventuell auch in der Nähe spezieller Infrastruktur für den Ölkürbisbau befindet. Der Effekt der Lage im g.g.A.-Gebiet ist sowohl in Niederösterreich als auch in der Steiermark und dem Burgenland positiv und auf dem 0,1%-Niveau signifikant. Die Nähe zu spezieller Infrastruktur wirkt auch positiv auf den geschätzten Ölkürbisanteil, ist aber nur mehr auf dem 5%-Niveau signifikant. Ein möglicher Grund dafür könnte sein, dass für die Standorte von Ölmühlen und Wasch- bzw. Trocknungsanlagen nur lückenhafte Daten verfügbar sind. So weist die räumliche Analyse der Residuen des SAR-Tobit-Modells in Abschnitt 4.3 auf einige Regionen hin, in denen der Ölkürbisanteil unterschätzt wird. Das könnte z. B. auf fehlende Standortdaten von spezieller Infrastruktur hindeuten. Solche Tendenzen sind etwa im Gebiet nordwestlich von Graz, in Teilen des Marchfelds sowie im Bezirk Melk und dem äußersten östlichen Teil des g.g.A.-Gebietes in Niederösterreich erkennbar.

Eine Einschränkung der Dummy-Variablen ist, dass sie nur die Differenz des geschätzten Ölkürbisanteils im Vergleich zur Basiskategorie beschreiben können (Lage außerhalb des g.g.A.-Gebietes und keine Nähe zu spezieller Infrastruktur). Eine Möglichkeit, um detailliertere Aussagen über den Effekt des g.g.A.-Gebietes auf den Ölkürbisbau machen zu können, wäre die Bildung von Interaktionstermen der beiden g.g.A.-Dummies mit den restlichen unabhängigen Variablen. So könnte man z. B. untersuchen, ob sich die Effekte der

anderen unabhängigen Variablen auf den Ölkürbisanteil in- und außerhalb des g.g.A.-Gebietes unterscheiden.

Im Verlauf der vorliegenden Arbeit wurde versuchsshalber ein SAR-Tobit-Modell gebildet, das für jede unabhängige Variable einen Interaktionsterm mit den g.g.A.-Dummies beinhaltet. So wurde im Prinzip ein eigenes Regressionsmodell für Gemeinden außerhalb des g.g.A.-Gebiets, im g.g.A.-Gebiet in Niederösterreich sowie im g.g.A.-Gebiet in der Steiermark und dem Burgenland in einem Modell vereint gebildet. Nur der räumliche Parameter  $Rho$  war für alle Gemeinden gleich. Der Fit dieses Modells war im Vergleich zum SAR-Tobit-Modell ohne Interaktionsterme geringfügig besser. Die Regressionskoeffizienten der signifikanten Variablen der unterschiedlichen Dummy-Kategorien unterschieden sich zwar etwas voneinander, ihre Vorzeichen blieben jedoch überwiegend unverändert. Außerdem verdreifachte sich so die Anzahl der zu schätzenden Parameter. Dadurch gestaltete sich die Interpretation der marginalen Effekte um einiges komplexer. Da die Ergebnisse der beiden Modelle sich von ihrer Grundaussage nicht wesentlich unterschieden, wurde zugunsten einer simpleren Modellinterpretation auf Interaktionsterme in den Regressionsmodellen verzichtet.

Die uneinheitlichen Aussagen der oben genannten Studien im Bezug auf die Wahl eines räumlichen Regressionsmodells (SAR- oder SEM-Modell) zeigen, dass es keine einheitliche Tendenz zu einem der beiden Modelle gibt. Es hängt vielmehr von der Problemstellung ab, von welcher Art der räumlichen Autokorrelation man ausgeht und welches Modell vorzuziehen ist.

Ein grundsätzliches Problem, in Bezug auf die räumlichen Effekte, stellt die Definition von Nachbarn und der Nähe zu spezieller Infrastruktur dar. Die Nachbarschaftsdefinitionen basieren in ähnlichen Anwendungsfällen auf direkter Nachbarschaft oder einem Entfernungsschwellenwert (siehe etwa SCHOLZ et al. 2013, ODGAARD et al. 2011 oder LIPPERT et al. 2012). In der vorliegenden Arbeit ist die Nachbarschaftsmatrix mit einem Entfernungsschwellenwert von 10,8 km definiert und auf Basis von inverser Distanz gewichtet. Obwohl diese Definition einen mit zunehmender Entfernung abnehmenden Effekt erlaubt, werden topografische Hindernisse wie Berge und Flüsse sowie Verkehrsverbindungen ignoriert. Für zukünftige Analysen könnte die Einbindung derartiger Informationen mit Hilfe geographischer Informationssysteme (GIS) eine realistischere Abbildung räumlicher Beziehungen (etwa auf Basis durchschnittlicher Fahrzeiten) schaffen.

Sowohl die Ergebnisse des Lagrange-Multiplier-Tests als auch des Moran's I-Tests sind auf dem 0,1%-Niveau signifikant. Das deutet auf räumliche Autokorrelation auf Gemeindeebene hin. Da die Teststatistik  $RLMlag$  signifikanter als  $RLMerr$  ist, wird von räumlicher Autokorrelation in der abhängigen Variable ausgegangen. Der Parameter  $Rho$  im SAR-Tobit-

Modell beträgt 0,67 und ist auf dem 0,1%-Niveau signifikant, was auf positive Agglomerationseffekte auf Gemeindeebene hindeutet.

Das gewählte SAR-Tobit-Modell erscheint auch aus theoretischer Sicht geeignet für die Problemstellung. Vor allem die hohen einzelbetrieblichen Maschinenkosten sowie die räumliche Bindung zu spezieller Infrastruktur könnten zu positiven Agglomerationseffekten im Ölkürbisbau führen.

Auch wenn die Testergebnisse hinsichtlich räumlicher Autokorrelation eher auf ein SAR-Tobit-Modell hindeuten, so ist aufgrund der fehlenden Standortdaten zu spezieller Infrastruktur trotzdem auch räumliche Autokorrelation der Fehler denkbar.

Um diese Annahme zu überprüfen, erscheint das von SCHOLZ et al. (2013) verwendete heteroskedastierobuste SARAR-Modell ein guter Ansatz. Dieses Modell ist in R derzeit noch nicht als Tobit-Spezifikation implementiert und konnte daher auch nicht in dieser Arbeit verwendet werden. Mit einem heteroskedastierobusten SARAR-Tobit-Modell, könnte man sowohl die noch immer vorhandene Heteroskedastizität als auch potentielle Autokorrelation der Fehler berücksichtigen. BIVAND und BRUNSTAD (2006) beschreiben besonders ausführlich, wie fehlende unabhängige Variablen (Fehlspezifikation) und Heteroskedastizität, zum Auftreten artifizieller räumlicher Autokorrelation führen können. Eine Möglichkeit, diesen Sachverhalt zu überprüfen wäre, zusätzliche unabhängige Variablen in das Modell einzufügen, welche die Erfüllung der Modellannahmen verbessern (z.B. zusätzliche Standorte spezieller Infrastruktur oder Bodenzahl). Ein erneuter Test auf räumliche Autokorrelation würde dann zeigen, ob noch immer räumliche Autokorrelation vorhanden ist, bzw. ob sich dadurch die Entscheidung hinsichtlich der Modellbildung (SAR- SEM oder SARAR-Modell) ändert.

DONFOUET et al. (2013, 18) weisen darauf hin, dass Regressionsmodelle, die mit dem MCMC-Verfahren geschätzt werden, robuster gegenüber Heteroskedastizität sind. Aus ihren Ausführungen ist aber nicht klar ersichtlich, inwieweit diese Annahme auch noch bei starker Heteroskedastizität (wie im vorliegenden SAR-Tobit-Modell) hält.

Die bereits angesprochene Erweiterung des Modells mit Paneldaten würde schließlich dazu beitragen, die Abhängigkeit der Ölkürbisanteile von den Bestimmungsfaktoren in einem größeren zeitlichen Kontext betrachten zu können. Sofern die Softwareverfügbarkeit in naher Zukunft gegeben ist, könnten so einige, in dieser Arbeit noch offen gebliebene, Fragen geklärt werden.



### **5.3 Weiterer Forschungsbedarf**

Die Ergebnisse dieser Arbeit liefern erste Erkenntnisse über die Bestimmungsfaktoren des Ölkürbisanbaus in Österreich. Es verbleiben aber noch einige ungelöste Probleme, die weiterer Forschungen bedürfen. Wie der RESET- und Breusch-Pagan-Test zeigen, liegen im SAR-Tobit-Modell noch immer leichte Fehlspezifikation und starke Heteroskedastizität vor. Außerdem umfasst die Analyse lediglich das Jahr 2011. Ein potentieller weiterer Schritt wäre, ein heteroskedastierobustes SARAR-Tobit-Modell mit Paneldaten und zusätzlichen Bestimmungsfaktoren zu bilden (z. B. Bodenzahl, weitere Standortdaten von spezieller Infrastruktur, Preisdaten von Ölkürbiskernen und Konkurrenzkulturen).

Weiterer Forschungsbedarf besteht auch bei der Berechnung von ADI, All und ATI für Dummy-Variablen in einem räumlichen Tobit-Modell. In dieser Hinsicht konnte im Verlauf der Arbeit noch kein geeignetes Berechnungsverfahren ausfindig gemacht werden.

Um lokale Potentiale für den Ölkürbisanbau eindeutig identifizieren zu können, bedarf es auch qualitativer Forschung vor Ort. Die Ergebnisse der räumlichen Residuenanalyse in Abschnitt 4.3 zeigen etwa, dass das niederösterreichische oder das burgenländische g.g.A.-Gebiet mögliche Ausgangspunkte für eine qualitative Untersuchung sein könnten. Ein Beispiel dafür, wie eine solche Untersuchung aussehen könnte, ist die Arbeit von SCHWARZ (2008). Sie untersucht im Rahmen einer qualitativen Fallstudie in 5 steirischen Bezirken den Beitrag der geographisch geschützten Angabe des Ölkürbisanbaus in Österreich zur ländlichen Entwicklung.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass die Berücksichtigung räumlicher Effekte für korrekte Modellergebnisse von zentraler Bedeutung ist. Darüber hinaus erweitern räumliche Modelle die Möglichkeiten der Ergebnisinterpretation, was zu einer vielschichtigen und somit realitätsnäheren Problemanalyse beiträgt. Sie stellen somit ein geeignetes Analyeinstrument zur Behandlung komplexer raumbezogener Fragestellungen dar. Nicht zuletzt soll mit dieser Arbeit daher auch zur stärkeren Auseinandersetzung mit räumlichen Tobit-Modellen in der Agrarökonomie angeregt werden.

## Literatur

- Agrarmarkt Austria, s.a.. ÖPUL 2007 - Integrierte Produktion Gemüse. Verfügbar in: [http://www.ama.at/Portal.Node/public?genetics.rm=PCP&genetics.pm=gti\\_full&p.contentid=10008.47303&MEBIPG.pdf](http://www.ama.at/Portal.Node/public?genetics.rm=PCP&genetics.pm=gti_full&p.contentid=10008.47303&MEBIPG.pdf) [Abfrage am 23.02.2013].
- Aguiar, A.P.D., Câmara, G., Escada, M.I.S., 2007. Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. *Ecological Modelling* 209, 169–188.
- Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N. (Hrsg.), Csaki, F. (Hrsg.). 2nd International Symposium on Information Theory. Budapest: Akademiai Kiado. (267-281).
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- Anselin, L., 2002. Under the hood Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27, 247–267.
- Anselin, L., 2003. An Introduction to Spatial Regression Analysis in R. Verfügbar in: [http://geodacenter.asu.edu/drupal\\_files/spdepintro.pdf](http://geodacenter.asu.edu/drupal_files/spdepintro.pdf) [Abfrage am 16.03.2014].
- Anselin, L., Bongiovanni, R., Lowenberg-DeBoer, J., 2004. A Spatial Econometric Approach to the Economics of Site-Specific Nitrogen Management in Corn Production. *American Journal of Agricultural Economics* 86, 675–687.
- Anselin, L., 2007. *Spatial Regression Analysis in R – A Workbook*. Urbana
- Autant-Bernard, C., LeSage, J., P. 2009. Quantifying Knowledge Spillovers using Spatial Econometric Models. *Journal of Regional Science* 51, 471–496.
- Backhaus, K., 2011. *Multivariate Analysemethoden eine anwendungsorientierte Einführung*. Springer, Berlin [u.a.].
- Bartus, T., 2005. Estimation of marginal effects using `margeff`. *The Stata Journal*, 3 (5), 309-329.
- Bichler, B., 2006. *Die möglichen Bestimmungsgründe der räumlichen Verteilung des ökologischen Landbaus in Deutschland*. Logos, Berlin.
- Bivand, R., Brunstad, R., 2006. Regional growth in Western Europe: detecting spatial misspecification using the R environment. *Papers in Regional Science* 85, 277–297.

- Bivand, R., 2013. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-68. Verfügbar in: <http://CRAN.R-project.org/package=spdep> [Abfrage am 30.11.2013].
- BMLFUW, s.a.. Steirisches Kürbiskernöl g.g.A. Verfügbar in: [http://www.lebensministerium.at/dms/lmat/lebensmittel/trad-lebensmittel/oel/steir\\_kuebiskernoel/Steirisches-Kuebiskernoel-ggA-d/Steirisches%20Kuebiskernoel%20ggA%20d.pdf](http://www.lebensministerium.at/dms/lmat/lebensmittel/trad-lebensmittel/oel/steir_kuebiskernoel/Steirisches-Kuebiskernoel-ggA-d/Steirisches%20Kuebiskernoel%20ggA%20d.pdf) [Abfrage am 23.04.2013].
- BMLFUW, 2001-2012. Grüne Berichte für die Jahre 2000-2011. Verfügbar in: <http://www.gruenerbericht.at/cm3/download/viewcategory/82-gruener-bericht-oesterreich.html> [Abfrage am 15.03.2013].
- BMLFUW, 2008. Deckungsbeiträge und Daten für die Betriebsplanung 2008. Eigenverlag, Wien
- BMLFUW, 2012. Gemeindedatenbank.
- Brady, M., Irwin, E., 2011. Accounting for Spatial Effects in Economic Models of Land Use: Recent Developments and Challenges Ahead. *Environmental and Resource Economics* 48, 487–509.
- Brandstetter, A., 2013. (Landwirtschaftskammer Niederösterreich). Mündliche Mitteilung vom 27.02.2013.
- Chatterjee, S., Hadi, A. S. 1988. *Sensitivity Analysis in Linear Regression*. John Wiley & Sons.
- Cretnik, A., 2013. (Geschäftsführer Gemeinschaft Steirisches Kürbiskernöl g.g.A). Schriftliche Mitteilung per E-Mail vom 21.02.2013.
- Diepenbrock, W., 1999. *Spezieller Pflanzenbau: 92 Tabellen*. Ulmer, Stuttgart.
- Donfouet, H., Jeanty, W. P., Malin, E., 2013. Accounting for Spatial Interactions in the Demand for Community-Based Health Insurance: A Bayesian Spatial Tobit Analysis. *Population Association of America 2013 Annual Meeting – Session 35: Insurance and Health Care Access*. Verfügbar in: <http://paa2013.princeton.edu/papers/130913> [Abfrage am 05.12.2013].
- Dörfler, M. 2011. *Ausgewählte Pflanzenschutzprobleme bei Ölkürbis im Bezirk Melk/NÖ in den Jahren 2009 – 2010*. Diplomarbeit. Universität für Bodenkultur Wien.
- ESRI, 2011. *ArcGIS Desktop: Version 10*. Redlands, CA: Environmental Systems Research Institute.

- Fahrmeir, L., Kneib, T., Lang, S., 2009. Regression: Modelle, Methoden und Anwendungen. Springer, Berlin; Heidelberg.
- Fingleton, B., López-Bazo, E., 2006. Empirical growth models with spatial effects. *Papers in Regional Science* 85, 177–198.
- Gellrich, M., Zimmermann, N.E., 2007. Investigating the regional-scale pattern of agricultural land abandonment in the Swiss mountains: A spatial statistical modelling approach. *Landscape and Urban Planning* 79, 65–76.
- Griffith, D.A., 2008. A comparison of four model specifications for describing small heterogeneous space-time datasets: Sugar cane production in Puerto Rico, 1958/59-1973/74\*. *Papers in Regional Science* 87, 341–355.
- Henningsen, A., 2013. censReg: Censored Regression (Tobit) Models. R-Paket Version 0.5-20. Verfügbar in: <http://CRAN.R-project.org/package=censReg> [Abfrage am 30.11.2013].
- Heyland, K.-U., Hanus, H., Keller, E.R., 2006. Ölfrüchte, Faserpflanzen, Arzneipflanzen und Sonderkulturen. Ulmer, Stuttgart.
- Kleiber, C., Zeileis, A., 2008. Applied Econometrics with R. Verfügbar in: <http://CRAN.R-project.org/package=AER> [Abfrage am 30.11.2013].
- Klotz, S., 2004. Cross Sectional Dependence in Spatial Econometric Models: With an Application to German Start-up Activity Data. LIT Verlag, Münster.
- Kulmer, G., Weber, J., 1990. Der Kürbis: Anbau, Verarbeitung. Stocker, Graz; Stuttgart.
- Landwirtschaftskammer Niederösterreich, s.a. Deckungsbeitragsvergleiche zwischen Betrieben und unterschiedlichen Kulturen auf Basis von Arbeitskreisergebnissen. s.l. Verfügbar in: [http://www.oekosozial.at/uploads/tx\\_osfopage/PPT\\_Schally.pdf](http://www.oekosozial.at/uploads/tx_osfopage/PPT_Schally.pdf) [Abfrage am 24.09.2013].
- Langyintuo, A., Mekuria, M., 2008 Assessing the influence of neighborhood effects on the adoption of improved agricultural technologies in developing agriculture. *African Journal of Agricultural and Resource Economics* 2 (2), 151-169.
- LeSage, J.P., Kelley, Pace R.K., 2009. Introduction to spatial econometrics, Statistics, textbooks and monographs. CRC Press, Boca Raton.
- Lippert, C., Chatzopoulos, T., Schmidtner, E., Aurbacher, J., 2012. Explaining the climate-dependent Distribution of Crops in Space - the Example of Corn and Corn-Cob-Mix in Baden-Württemberg. Presented at the Schriften der GEWISOLA, Halle, pp. 91–102.

- Long, J.S., 1997. Regression models for categorical and limited dependent variables, Advanced quantitative techniques in the social sciences. Sage Publications, Thousand Oaks.
- Maschinenring, 2010. Maschinenring aktuell, Ausgabe 27 – September 2010. Zwettl-Weitra, Waldviertel Nord, Manhartsberg, Krems-Gföhl. Verfügbar in: <http://www.dolne-rakusko.info/magazin/00/artikel/71726/doc/d/MR%20Aktuell%20September%202010.pdf?ok=j> [Abfrage am 12.08.2013].
- Meyer-Aurich, A., Griffin, T.W., Herbst, R., Giebel, A., Muhammad, N., 2010. Spatial econometric analysis of a field-scale site-specific nitrogen fertilizer experiment on wheat (*Triticum aestivum* L.) yield and quality. *Computers and Electronics in Agriculture* 74, 73–79.
- Neubauer, V., 2009. Ertragsversuche bei Ölkürbissorten in Niederösterreich. Diplomarbeit. Universität für Bodenkultur Wien.
- Odgaard, M.V., Bøcher, P.K., Dalgaard, T., Svenning, J.-C., 2011. Climatic and non-climatic drivers of spatiotemporal maize-area dynamics across the northern limit for maize production—A case study from Denmark. *Agriculture, Ecosystems & Environment* 142, 201–302.
- Plant, R. E., 2012. Spatial data analysis in ecology and agriculture using R. CRC Press, Boca Raton.
- R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Wien, Österreich. Verfügbar in: <http://www.R-project.org/> [Abfrage am 25.11.2013].
- RStudio, 2013. RStudio: Integrierte Entwicklungsumgebung für R (Version 0.97.551) [Computer software]. Boston, MA. Verfügbar in: <http://www.rstudio.org/> [Abfrage am 25.11.2013].
- Ripley, B., Hornik, K., Gebhardt, A. und Firth, D. 2012. MASS. Support Functions and Datasets for Venables and Ripley's MASS. R-Paket Version 7.3-17. Verfügbar in: <http://www.stats.ox.ac.uk/pub/MASS4/> [Abfrage am 06.11.2013].
- Sachs, L., Hedderich, J. 2009. Angewandte Statistik. Methodensammlung mit R. Dreizehnte, aktualisierte und erweiterte Auflage. Springer Verlag Berlin, Berlin.

- Scholz, L., Meyer-Aurich, A., Kirschke, D., 2013. Bestimmungsfaktoren der Silomaisproduktion – Eine räumlich-ökonomische Analyse. In: Bahrs, Becker, Birner, Brockmeier, Dabbert, Doluschitz, Grethe, Lippert, Thiele, Hrsg. Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e. V. Band 48. Münster: Landwirtschaftsverlag GmbH.
- Schwarz, S., 2008. Steirisches Kübiskernöl: Beitrag der 'geographisch gesegneten Angabe' zur ländlichen Entwicklung. Diplomarbeit. Universität für Bodenkultur Wien.
- Schuster, W., 1977. Der Ölkürbis (*Cucurbita pepo* L.): e. monograph. Darst, Fortschritte im Acker- und Pflanzenbau. Parey, Berlin ; Hamburg.
- Sperber, J., 1988. Öl -und Eiweisspflanzen: Anbau, Kultur, Ernte. Österreichischer Agrarverlag, Wien.
- Umweltbundesamt GmbH, 2006. CORINE-Landbedeckung 2006. Verfügbar in: [http://www.umweltbundesamt.at/umweltsituation/umweltinfo/opendata/oed\\_landbedeckung/?cgiproxy\\_url=http%3A%2F%2Fhttpapp5.umweltbundesamt.at%2Fdata%2Fdataset.jsf%3Bjsessionid%3D29B8DB8E467B7EBB76EC515EB343C36A%3Fid%3D1](http://www.umweltbundesamt.at/umweltsituation/umweltinfo/opendata/oed_landbedeckung/?cgiproxy_url=http%3A%2F%2Fhttpapp5.umweltbundesamt.at%2Fdata%2Fdataset.jsf%3Bjsessionid%3D29B8DB8E467B7EBB76EC515EB343C36A%3Fid%3D1) [Abfrage am 05.03.2013].
- Wilhelm, S. und Godinho de Matos M., 2013. spatialprobit: Spatial Probit Models. R Paket Version 0.9-9. Verfügbar in: <http://CRAN.R-project.org/package=satialprobit> [Abfrage am 30.11.2013].
- Wooldridge, J.M., 2009. Introductory econometrics: a modern approach. South Western, Cengage Learning, Mason, OH.
- Yee, T., W., 2013. VGAM: Vector Generalized Linear and Additive Models. R Paket Version 0.9-3. Verfügbar in: <http://CRAN.R-project.org/package=VGAM> [Abfrage am 30.11.2013].
- Zeileis, A., und Hothorn T., 2013. lmerTest: Testing linear Regression Models. R package version 0.9-32. Verfügbar in: <http://cran.r-project.org/web/packages/lmerTest/index.html> [Abfrage am 30.11.2013].
- Zentralanstalt für Meteorologie und Geodynamik 2013. Langjährige Klimamittelwerte von 1971-2000 für Jahresdurchschnittstemperatur, Anzahl der Frostwechsellage sowie Sommerniederschlag (von April bis September) mit der Auflösung 250 m x 250 m. Nach persönlicher Anfrage übermittelt.

# Anhang

## A. Detaillierte Ergebnistabellen der Regressionsmodelle

Tabelle 6: Ergebnisse OLS-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$   |          |                |         |           |
|---|----------|----------------|---------|-----------|
| Anzahl der Beobachtungen: 1.347 Gemeinden               |          |                |         |           |
| Unabhängige Variablen                                   | Wert     | Standardfehler | t-Wert  | p-Wert    |
| Konstante   | 0,0893   | 0,0273         | 3,2711  | 0,0011    |
| g.g.A.-Gebiet Nö.                                       | 0,8578   | 0,0778         | 11,0318 | < 2,2E-16 |
| g.g.A.-Gebiet Stmk.-Bgld.                               | 1,3199   | 0,0762         | 17,3221 | < 2,2E-16 |
| Nähe zu spezieller Infrastruktur                        | 0,5496   | 0,0786         | 6,9928  | 4,2E-12   |
| Temperatur <sub>zentr</sub> $\times 10$                 | 0,0158   | 0,0038         | 4,1524  | 3,5E-05   |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup> | 0,0007   | 0,0003         | 2,4780  | 0,0133    |
| Niederschlag <sub>zentr</sub> / 10                      | 0,0077   | 0,0017         | 4,3900  | 1,2E-05   |
| GVE pro ha LF <sub>zentr</sub> $\times 10$              | -0,0239  | 0,0050         | -4,8070 | 1,7E-06   |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$       | 0,1677   | 0,0227         | 7,3823  | 2,7E-13   |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$   | 0,0529   | 0,0139         | 3,8143  | 0,0001    |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$            | 0,0749   | 0,0206         | 3,6302  | 0,0003    |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$         | -0,2405  | 0,0470         | -5,1139 | 3,6E-07   |
| Gütemaße  |          |                |         |           |
| R <sup>2</sup>  | 0,71     |                |         |           |
| AIC   | 2.327,71 |                |         |           |
| Teststatistiken   |          |                |         |           |
| RESET-Test  | 95,54    |                |         | < 2,2E-16 |
| Breusch-Pagan Test                                      | 275,70   |                |         | < 2,2E-16 |
| RLMlag  | 107,39   |                |         | < 2,2E-16 |
| RLMerr  | 29,58    |                |         | 5,3E-08   |
| Moran's-I Residuen                                      | 0,35     |                |         | < 2,2E-16 |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\* : > 0,001 – 0,01 (1%-Niveau); \* : > 0,01 – 0,05 (5%-Niveau)

Tabelle 7: Marginale Effekte OLS-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$<br>N = 1347 Gemeinden |                   |             |
|---|-------------------|-------------|
| Unabhängige Variablen   | Marginale Effekte | Signifikanz |
| g.g.A.-Gebiet Nö.   | 85,78             | ***         |
| g.g.A.-Gebiet Stmk.-Bgld.   | 131,99            | ***         |
| Nähe zu spezieller Infrastruktur  | 54,96             | ***         |
| Temperatur <sub>zentr</sub> $\times 10$                                     | 1,58              | ***         |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup>                     | 0,07              | ***         |
| Niederschlag <sub>zentr</sub> / 10  | 0,77              | **          |
| GVE pro ha LF <sub>zentr</sub> $\times 10$                                  | -2,39             | *           |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$                           | -0,02             | ***         |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$                       | 0,17              | ***         |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$                                | 0,05              | .           |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$                             | 0,08              | ***         |

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\*: > 0,001 – 0,01 (1%-Niveau); \*: > 0,01 – 0,05 (5%-Niveau)

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)



Tabelle 8: Ergebnisse Tobit-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$      |          |                |          |           |
|--|----------|----------------|----------|-----------|
| Anzahl der Beobachtungen: 1.347 Gemeinden                  |          |                |          |           |
| Unabhängige Variablen                                      | Wert     | Standardfehler | t-Wert   | p-Wert    |
| Konstante  | -0,7868  | 0,0669         | -11,7680 | 5,7E-32   |
| g.g.A.-Gebiet Nö.  | 1,4690   | 0,1052         | 13,9614  | < 2,2E-16 |
| g.g.A.-Gebiet Stmk.-Bgld.                                  | 2,1258   | 0,0951         | 22,3413  | < 2,2E-16 |
| Nähe zu spezieller Infrastruktur                           | 0,4799   | 0,0941         | 5,0988   | 3,4E-07   |
| Temperatur <sub>zentr</sub> <sup>x</sup> 10                | 0,0234   | 0,0067         | 3,5088   | 0,0005    |
| (Temperatur <sub>zentr</sub> <sup>x</sup> 10) <sup>2</sup> | 0,0015   | 0,0005         | 3,0767   | 0,0021    |
| Niederschlag <sub>zentr</sub> / 10                         | 0,0104   | 0,0049         | 2,1230   | 0,0338    |
| GVE pro ha LF <sub>zentr</sub> <sup>x</sup> 10             | -0,0542  | 0,0093         | -5,8183  | 5,9E-09   |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$          | 0,2824   | 0,0356         | 7,9316   | 2,2E-15   |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$      | 0,0430   | 0,0238         | 1,8090   | 0,0705    |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$               | 0,2262   | 0,0365         | 6,1945   | 5,8E-10   |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$            | -0,3659  | 0,0840         | -4,3578  | 1,3E-05   |
| Log Sigma  | -0,1224  | 0,0295         | -4,1467  | 3,4E-05   |
| <b>Gütemaße</b>  |          |                |          |           |
| R <sup>2</sup> <sup>b)</sup>                               | 0,73     |                |          |           |
| AIC  | 2.290,10 |                |          |           |
| <b>Teststatistiken</b>                                     |          |                |          |           |
| RESET-Test   | 6,34     |                |          | 0,0421    |
| Breusch-Pagan Test   | 96,24    |                |          | < 2,2E-16 |
| RLMlag   | / a)     |                |          |           |
| RLMerr   | / a)     |                |          |           |
| Moran's-I Residuen   | 0,54     |                |          | < 2,2E-16 |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010)

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\* : > 0,001 – 0,01 (1%-Niveau); \* : > 0,01 – 0,05 (5%-Niveau)

a) Für Felder, die mit „/a“ gekennzeichnet sind, konnte kein Wert berechnet werden

b) Das R<sup>2</sup> wurde als quadriertes Korrelationskoeffizient von  $\hat{y}$  und y berechnet (siehe Wooldridge 2009, 592).

**Tabelle 9: Zusammenfassung unterschiedlicher Nachbarschaftsmatrizen, basierend auf kontinuierlicher Nachbarschaft bzw. Schwellenentfernungen mit inverser Distanzgewichtung**

|                             | <b>N</b>     | <b>Anzahl<br/>Verbindungen</b> | <b>Min.</b> | <b>Max</b> | <b>Mittelwert</b> |
|-----------------------------|--------------|--------------------------------|-------------|------------|-------------------|
| W.cont.                     | 1.347        | 7.256                          | 1           | 13         | 5,4               |
| <b>W.min1<br/>(10,8 km)</b> | <b>1.347</b> | <b>18.710</b>                  | <b>1</b>    | <b>33</b>  | <b>13,9</b>       |
| W.15km                      | 1.347        | 34.934                         | 1           | 56         | 25,9              |
| W.20km                      | 1.347        | 59.274                         | 4           | 98         | 44,0              |

Quelle: Eigene Darstellung

Tabelle 10: Ergebnisse SAR-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$   |          |                |          |           |
|---|----------|----------------|----------|-----------|
| Anzahl der Beobachtungen: 1.347 Gemeinden               |          |                |          |           |
| Unabhängige Variablen                                   | Wert     | Standardfehler | z-Wert   | p-Wert    |
| Konstante   | 0,0241   | 0,02326        | 1,03766  | 0,2994    |
| g.g.A.-Gebiet Nö.                                       | 0,2504   | 0,04957        | 5,05094  | 4,4E-07   |
| g.g.A.-Gebiet Stmk.-Bgld.                               | 0,2891   | 0,0524         | 5,51698  | 3,5E-08   |
| Nähe zu spezieller Infrastruktur                        | 0,2011   | 0,04687        | 4,29036  | 1,8E-05   |
| Temperatur <sub>zentr</sub> $\times 10$                 | 0,0034   | 0,00281        | 1,19621  | 0,2316    |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup> | 0,0003   | 0,00018        | 1,46648  | 0,1425    |
| Niederschlag <sub>zentr</sub> / 10                      | 0,0051   | 0,00181        | 2,82428  | 0,0047    |
| GVE pro ha LF <sub>zentr</sub> $\times 10$              | -0,0194  | 0,00391        | -4,97781 | 6,4E-07   |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$       | 0,0838   | 0,01542        | 5,43629  | 5,4E-08   |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$   | 0,0130   | 0,00975        | 1,33056  | 0,1833    |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$            | 0,0576   | 0,01445        | 3,98713  | 6,7E-05   |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$         | -0,1139  | 0,03476        | -3,27632 | 0,0011    |
| Rho   | 0,7230   | / a)           | / a)     | < 2,2E-16 |
| <b>Gütemaße</b>   |          |                |          |           |
| R <sup>2</sup> <sup>b)</sup>                            | 0,82     |                |          |           |
| AIC   | 1.782,94 |                |          |           |
| <b>Teststatistiken</b>                                  |          |                |          |           |
| RESET-Test  | 205,63   |                |          | < 2,2E-16 |
| Breusch-Pagan Test                                      | 196,38   |                |          | < 2,2E-16 |
| Moran's-I Residuen                                      | 0,01     |                |          | 0,2473    |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010).

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\* : > 0,001 – 0,01 (1%-Niveau); \* : > 0,01 – 0,05 (5%-Niveau)

a) Für Felder, die mit „/“ gekennzeichnet sind, konnte kein Wert berechnet werden.

b) Das R<sup>2</sup> wurde als quadriertes Korrelationskoeffizient von  $\hat{y}$  und  $y$  berechnet (siehe Wooldridge 2009, 592).

Tabelle 11: Marginale Effekte SAR-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$<br>N = 1347 Gemeinden |       |       |        |             |
|---|-------|-------|--------|-------------|
| Unabhängige Variablen   | ADI   | All   | ATI    | Signifikanz |
| g.g.A.-Gebiet Nö.   | 27,90 | 64,83 | 92,73  | ***         |
| g.g.A.-Gebiet Stmk.-Bgld.   | 32,22 | 74,86 | 107,07 | ***         |
| Nähe zu spezieller Infrastruktur  | 22,41 | 52,07 | 74,47  | ***         |
| Temperatur <sub>zentr</sub> $\times 10$                                     | 0,38  | 0,87  | 1,25   |             |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup>                     | 0,03  | 0,07  | 0,10   |             |
| Niederschlag <sub>zentr</sub> / 10  | 0,57  | 1,33  | 1,90   | **          |
| GVE pro ha LF <sub>zentr</sub> $\times 10$                                  | -2,17 | -5,04 | -7,20  | ***         |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$                           | 0,09  | 0,22  | 0,31   | ***         |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$                       | 0,01  | 0,03  | 0,05   |             |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$                                | 0,06  | 0,15  | 0,21   | ***         |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$                             | -0,13 | -0,29 | -0,42  | **          |

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\*: > 0,001 – 0,01 (1%-Niveau); \*: > 0,01 – 0,05 (5%-Niveau)

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010).

Tabelle 12: Ergebnisse SAR-Tobit-Modell

| Abhängige Variable = $\ln(1 + \text{Ölkürbisanteil})$   |          |                    |         |           |
|---|----------|--------------------|---------|-----------|
| Anzahl der Beobachtungen: 1.347 Gemeinden               |          |                    |         |           |
| Unabhängige Variablen                                   | Wert     | Standardabweichung | t-Wert  | p-Wert    |
| Konstante   | -0,3169  | 0,0584             | -5,4215 | 7,0E-08   |
| g.g.A.-Gebiet Nö.                                       | 0,5470   | 0,1033             | 5,2928  | 1,4E-07   |
| g.g.A.-Gebiet Stmk.-Bgld.                               | 0,7505   | 0,1058             | 7,0924  | 2,1E-12   |
| Nähe zu spezieller Infrastruktur                        | 0,1865   | 0,0898             | 2,0785  | 0,0379    |
| Temperatur <sub>zentr</sub> $\times 10$                 | 0,0051   | 0,0061             | 0,8387  | 0,4018    |
| (Temperatur <sub>zentr</sub> $\times 10$ ) <sup>2</sup> | 0,0008   | 0,0004             | 1,7657  | 0,0777    |
| Niederschlag <sub>zentr</sub> / 10                      | 0,0086   | 0,0041             | 2,1132  | 0,0348    |
| GVE pro ha LF <sub>zentr</sub> $\times 10$              | -0,0425  | 0,0081             | -5,2397 | 1,9E-07   |
| $\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}$       | 0,1649   | 0,0333             | 4,9545  | 8,2E-07   |
| $[\ln(1 + \text{Körnermaisanteil})_{\text{zentr}}]^2$   | -0,0113  | 0,0220             | -0,5121 | 0,6087    |
| $\ln(1 + \text{Ackerfläche})_{\text{zentr}}$            | 0,1933   | 0,0334             | 5,7878  | 8,9E-09   |
| $\ln(1 + \text{LF pro Betrieb})_{\text{zentr}}$         | -0,2177  | 0,0763             | -2,8541 | 0,0044    |
| Sigma   | 0,6603   | 0,0504             | 13,0973 | < 2,2E-16 |
| Rho   | 0,6667   | 0,0329             | 20,2888 | < 2,2E-16 |
| <b>Gütemaße</b>   |          |                    |         |           |
| R <sup>2</sup> <sup>a)</sup>                            | 0,82     |                    |         |           |
| AIC   | 1.952,26 |                    |         |           |
| <b>Teststatistiken</b>                                  |          |                    |         |           |
| RESET-Test  | 18,48 b) |                    |         | 0,0001    |
| Breusch-Pagan Test                                      | 73,67 b) |                    |         | 2,4E-11   |
| Moran's-I Residuen                                      | 0,02     |                    |         | 0,0513    |

Quelle: Eigene Berechnungen, Daten von BMLFUW s.a., BMLFUW 2012, ZAMG 2013, Gemeinschaft Steirisches Kürbiskernöl g.g.A. (2013) sowie Maschinenring (2010).

Signifikanzniveaus: \*\*\* : 0 - 0,001 (0,1%-Niveau); \*\* : > 0,001 – 0,01 (1%-Niveau); \* : > 0,01 – 0,05 (5%-Niveau)

a) Das R<sup>2</sup> wurde als quadrierter Korrelationskoeffizient von  $\hat{y}$  und  $y$  berechnet (siehe Wooldridge 2009, 592).

b) Der Reset- und Breusch-Pagan-Test des SAR-Tobit-Modells beruhen auf einer ML-Schätzung eines SAR-Modells mit der latenten Variable  $y^*$  des SAR-Tobit-Modells als abhängige Variable.

## B. R-Skript zur Berechnung der Regressionsmodelle

```
#####  
#####  
#####1) Simple OLS-Modell (N=1347)#####  
#####  
#####  
  
model.ols <- lm(OelkAntLog ~  
                DNSB + SI5  
                + I(MT * 10) + I((MT * 10)^2)  
                + I(MNs/10)  
                + I(MGVEproHa * 10)  
                + MKmAntLog + I(MKmAntLog^2)  
                + MAckerflLog  
                + MLNFBetrLog  
                ,data= gemeindedaten  
)  
  
#Zusammenfassung und Fit:  
summary(model.ols)  
AIC(model.ols)  
  
#Diagnostiktests:  
resettest(model.ols)  
bptest(model.ols)  
  
#Berechnen heteroskedastischer Standardfehler:  
coef.hetrob.ols <- coeftest(model.ols, vcov = vcovHC)  
coef.hetrob.ols  
  
#Berechnung der gefitteten Werte in Level-Form:  
res.ols <- resid(model.ols)  
fw.ols <- fitted(model.ols)  
summary(res.ols)  
summary(fw.ols)  
alpha.dach.ols <- 1/1347 * sum(exp(res.ols))  
fw.ols.level <- alpha.dach.ols * (exp(fw.ols)-1)  
summary(fw.ols.level)  
res.ols.level <- gemeindedaten$OelkAnt - fw.ols.level  
summary(res.ols.level)
```

```
#####
#####
#####2) Tobit-Modell (N=1347)#####
#####
#####

model.tobit <- censReg(OelkAntLog ~
  DNSB + SI5
  + I(MT * 10) + I((MT * 10)^2)
  + I(MNs/10)
  + I(MGVEproHa * 10)
  + MKmAntLog + I(MKmAntLog^2)
  + MAckerflLog
  + MLNFBetrLog
, data= gemeindedaten
)

#Summary Measures:
summary(model.tobit)
names(model.tobit)
AIC(model.tobit)

#####
###Händische Berechnung der marginal effects (APE):
#Bildung der notwendigen Variablen:
alle.parameter <- coef(model.tobit, logSigma =FALSE)
sigma.tobit <- alle.parameter["sigma"]
koef.tobit <- alle.parameter[!names(alle.parameter) %in% c("sigma")]
koef.tobit <- koef.tobit[-1]
length(koef.tobit)
attach(gemeindedaten@data)
X <- cbind(as.numeric(levels(DummyGGANOE))[DummyGGANOE],
  as.numeric(levels(DummyGGASTmkBglD))[DummyGGASTmkBglD],
  as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2), I(MNs/10),
  I(MGVEproHa*10), MKmAntLog, I(MKmAntLog^2), MAckerflLog, MLNFBetrLog)

summary(X)
dim(X)
colnames(X) <- c("ggA-NOE", "ggA-StmkBglD", "sI5km", "MT*10", "(MT*10)^2",
  "MNs/10", "MGVEproHa*10", "MKmAntLog", "MKmAntLog^2", "MAckerflLog",
  "MLNFBetrLog")
str(X)
detach(gemeindedaten@data)

#APE für kontinuierliche Variablen:
scaling.factor.APE.tobit <- 1/1347 * sum(pnorm(X%%koef.tobit/sigma.tobit))
scaling.factor.APE.tobit
APE.tobit <- koef.tobit * scaling.factor.APE.tobit
APE.tobit
str(APE.tobit)
APE.tobit <- cbind(as.data.frame(APE.tobit))
APE.tobit

#APE für die diskreten Variablen (Dummies):
#Erwartungswert schätzen, wenn je ein Dummy 0 ist und 1 ist, daraus die
#Differenz berechnen und die 1347 Werte danach aufsummieren, bevor sie
#abschließend durch 1347 dividiert werden
#(Details siehe wooldridge 2009, 592).
```

## Anhang

---

```
#Eigene Variablen bilden, für welche die Dummy Variablen auf 0 bzw. 1
  gesetzt werden.
gemeindedaten$Dummy.0 <- rep(0, times=length(gemeindedaten$Oe1kAntLog))
gemeindedaten$Dummy.1 <- rep(1, times=length(gemeindedaten$Oe1kAntLog))
#eigene X-Matrix definieren, bei der die Dummies nach der Reihe auf 0 bzw 1
  gesetzt werden:
attach(gemeindedaten@data)
#DNSB1:
X.DNSB1.0 <- cbind(a1, Dummy.0,
  as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld],
  as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2), I(MNs/10),
  I(MGVEproHa*10), MKmAntLog, I(MKmAmtLog^2), MAckerf1Log, MLNFBetrLog)

X.DNSB1.1 <- cbind(a1, Dummy.1,
  as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld],
  as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2), I(MNs/10),
  I(MGVEproHa*10), MKmAntLog, I(MKmAmtLog^2), MAckerf1Log, MLNFBetrLog)

#DNSB2:
X.DNSB2.0 <- cbind(a1, as.numeric(levels(DummyGGANOE))[DummyGGANOE],
  Dummy.0, as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2),
  I(MNs/10), I(MGVEproHa*10), MKmAntLog, I(MKmAmtLog^2), MAckerf1Log,
  MLNFBetrLog)

X.DNSB2.1 <- cbind(a1, as.numeric(levels(DummyGGANOE))[DummyGGANOE],
  Dummy.1, as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2),
  I(MNs/10), I(MGVEproHa*10), MKmAntLog, I(MKmAmtLog^2), MAckerf1Log,
  MLNFBetrLog)

#SI5:
X.SI5.0 <- cbind(a1, as.numeric(levels(DummyGGANOE))[DummyGGANOE],
  as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld], Dummy.0,
  I(MT*10), I((MT*10)^2), I(MNs/10), I(MGVEproHa*10), MKmAntLog,
  I(MKmAmtLog^2), MAckerf1Log, MLNFBetrLog)

X.SI5.1 <- cbind(a1, as.numeric(levels(DummyGGANOE))[DummyGGANOE],
  as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld], Dummy.1,
  I(MT*10), I((MT*10)^2), I(MNs/10), I(MGVEproHa*10), MKmAntLog,
  I(MKmAmtLog^2), MAckerf1Log, MLNFBetrLog)

detach(gemeindedaten@data)
summary(X.DNSB1.0)
summary(X.DNSB1.1)
summary(X.DNSB2.0)
summary(X.DNSB2.1)
summary(X.SI5.0)
summary(X.SI5.1)
summary(X)

#Umbenennen von koef.tobit in beta:
beta <- koef.tobit
length(beta)
beta
sigma <- sigma.tobit
```



```
#Formel für die Berechnung des Erwartungswertes von y:
#für DNSB1 = 0
E_y.DNSB1.0 <- pnorm(X.DNSB1.0 %% beta/sigma) * X.DNSB1.0 %% beta + sigma
  * dnorm(X.DNSB1.0 %% beta / sigma)
#für DNSB1 = 1
E_y.DNSB1.1 <- pnorm(X.DNSB1.1 %% beta/sigma) * X.DNSB1.1 %% beta + sigma
  * dnorm(X.DNSB1.1 %% beta / sigma)
#für DNSB2 = 0
E_y.DNSB2.0 <- pnorm(X.DNSB2.0 %% beta/sigma) * X.DNSB2.0 %% beta + sigma
  * dnorm(X.DNSB2.0 %% beta / sigma)
#für DNSB2 = 1
E_y.DNSB2.1 <- pnorm(X.DNSB2.1 %% beta/sigma) * X.DNSB2.1 %% beta + sigma
  * dnorm(X.DNSB2.1 %% beta / sigma)
#für SI5 = 0
E_y.SI5.0 <- pnorm(X.SI5.0 %% beta/sigma) * X.SI5.0 %% beta + sigma *
  dnorm(X.SI5.0 %% beta / sigma)
#für SI5 = 1
E_y.SI5.1 <- pnorm(X.SI5.1 %% beta/sigma) * X.SI5.1 %% beta + sigma *
  dnorm(X.SI5.1 %% beta / sigma)

#Differenzen bilden (E_y.dummy.1 - E_y.dummy.0)
#für DNSB1:
APE.DNSB1 <- 1/length(gemeindedaten$oe1kAntLog) * sum(E_y.DNSB1.1 -
  E_y.DNSB1.0)
APE.DNSB1
#für DNSB2
APE.DNSB2 <- 1/length(gemeindedaten$oe1kAntLog) * sum(E_y.DNSB2.1 -
  E_y.DNSB2.0)
APE.DNSB2
#für SI5:
APE.SI5 <- 1/length(gemeindedaten$oe1kAntLog) * sum(E_y.SI5.1 - E_y.SI5.0)
APE.SI5
APE.tobit["DNSB1",] <- APE.DNSB1
APE.tobit["DNSB2",] <- APE.DNSB2
APE.tobit["SI51",] <- APE.SI5
APE.tobit
#####

#####
#####3) Nachbarschaftsmatrizen und Tests auf räumliche Autokorrelation####
#####

#####3.1) Erstellung von Nachbarschaftsmatrizen:#####
#####
#####
#Notwendige Packages installieren:
#library(spdep)
#Koordinaten und Gemeinde-IDs extrahieren und als Matrizen abspeichern:
dim(gemeindedaten)
CoordsGesamt <- as.matrix(gemeindedaten@data[51:52])
GemIDGesamt <- as.matrix(gemeindedaten@data["G_Name_shp"])
summary(GemIDGesamt)
summary(CoordsGesamt)
dim(GemIDGesamt)
dim(CoordsGesamt)
```

```
#Um die höchste Entfernung zu bekommen, bei der alle noch Nachbarn sind
#wird in einem ersten Schritt eine nachbarschaftsliste erstellt, bei der
#jede Gemeinde zumindest eine Nachbargemeinde hat:
Oe1kNN1 <- knn2nb(knearneigh(CoordsGesamt, k=1), row.names=GemIDGesamt)

#Davon werden im zweiten Schritt die Distanzen berechnet:
Oe1kDistKNN1<-unlist(nbdists(Oe1kNN1, CoordsGesamt))
summary(Oe1kDistKNN1)

#Schließlich wird die Maximaldistanz separat abgespeichert, weil man sie
#später als Input für das Distanzband der Nachbarschaftsmatrix braucht.
MaxEntf <- max(Oe1kDistKNN1) #Speichern der max. Entf.
MaxEntf

#Mit der Funktion dnearneigh() werden anhand der Koordinaten und der vorhin
#gespeicherten Entfernungen Listen an Nachbarn basierend auf der jeweiligen
#Entfernung berechnet. Als Namen (IDS) für die Koordinaten können etwa die
#Gemeindenamen oder GKZ5 dienen.
NB.cont <- poly2nb(gemeindedaten, row.names=gemeindedaten@data[,
"G_Name_shp" ], queen=TRUE)
summary(NB.cont)
NB.min1<-dnearneigh(CoordsGesamt, d1=0, d2=MaxEntf, row.names=GemIDGesamt)

#Als Dist."NameListe" werden danach aufbauend auf die Distanzbasierte
Nachbarschaftsliste die Distanzen der einzelnen Gemeinden zueinander
berechnet:
Dist.NB.cont <- nbdists(NB.cont, CoordsGesamt)
summary(unlist(Dist.NB.cont))
Dist.NB.min1 <-nbdists(NB.min1, CoordsGesamt)
summary(unlist(Dist.NB.min1))

#Von der vorhin erstellten Listen werden im Anschluss die Kehrwerte der
#Distanzen berechnet um zu einer Inverse-Distance zu kommen:
InvDist.NB.cont <- lapply(Dist.NB.cont, function(x) 1/x)
InvDist.NB.min1 <- lapply(Dist.NB.min1, function(x) 1/x)

#Die hinzugefügte Gewichtung nach Inverse-Distanz wird als neue Liste
#abgespeichert (Bis auf die Liste mit den direkten Nachbarschaften)
W.ges.cont <- nb2listw(NB.cont, style="w", zero.policy=TRUE)
W.ges.cont.invdist <- nb2listw(NB.cont, glist=InvDist.NB.cont, style="w",
zero.policy=TRUE)
W.ges.min1 <- nb2listw(NB.min1, glist=InvDist.NB.min1, style="w")

#Sie kann in einer ersten Summary betrachtet werden.
summary(W.ges.cont, zero.policy=TRUE)
summary(W.ges.cont.invdist, zero.policy=TRUE)
summary(W.ges.min1)

#Und auch grafisch dargestellt werden.
plot(W.ges.cont, CoordsGesamt, main="Nachbarschaften bei direkter
Nachbarschaft")
plot(W.ges.cont.invdist, CoordsGesamt)
plot(W.ges.min1, CoordsGesamt)

#Schließlich kann die Nachbarschaftsliste mit Inverse-Distance Gewichtung
#in eine spatial weights Matrix umgewandelt werden (Je nachdem, für welches
#Regressionsmodell man sie später verwendet
```

```
Mat.w.ges.cont <- listw2mat(w.ges.cont)
Mat.w.ges.cont.invdist <- listw2mat(w.ges.cont.invdist)
Mat.w.ges.min1 <- listw2mat(w.ges.min1)

#Aufbau von Matrix anschauen:
dim(Mat.w.ges.cont)

#Umwandeln der Nachbarschaftsmatrix in eine sparse Matrix (ist für das
#Sartobit-Modell notwendig und dient der Reduktion der Rechenzeit).
Mat.w.ges.cont.sparse <- as(Mat.w.ges.cont, "sparseMatrix")
Mat.w.ges.cont.invdist.sparse <- as(Mat.w.ges.cont.invdist, "sparseMatrix")
dim(Mat.w.ges.cont.invdist.sparse)
Mat.w.ges.min1.sparse <- as(Mat.w.ges.min1, "sparseMatrix")

#####
#####
#####3.2) Tests auf räumliche Autokorrelation:#####
#####
#####

#####
#####3.2.1) OLS-Modell (N=1334)#####
#####

#Contiguity Neighbour (queen)
lm.LMtests(model.ols, listw=w.ges.cont, test="all")
lm.LMtests(model.ols, listw=w.ges.cont.invdist, test="all")
#Inverse-Distance: min1 (ca. 11 km)
summary(model.ols)
lm.LMtests(model.ols, listw=w.ges.min1, test="all")

#Das Resultat zeigt, dass sowohl für die kontinuierliche Nachbarschaft, als
#auch für die auf inverser Distanz beruhende Nachbarschaft bei min1
#(ca.11km) ein Spatial-Lag Modell vorzuziehen ist.
lm.morantest(model.ols, w.ges.cont, alternative="two.sided")
lm.morantest(model.ols, w.ges.cont.invdist, alternative="two.sided")
lm.morantest(model.ols, w.ges.min1, alternative="two.sided")

#####
#####
#####3.2.3) Tobit-Modell (N=1334)#####
#####

##Statistische Tests auf räumliche Autokorrelation
moran.test(res.tobit.uncensored, listw=w.ges.cont, randomisation=FALSE,
           alternative="two.sided")
moran.test(res.tobit.uncensored, listw=w.ges.cont.invdist,
           randomisation=FALSE, alternative="two.sided")
moran.test(res.tobit.uncensored, listw=w.ges.min1, randomisation=FALSE,
           alternative="two.sided")
#Die Analyse wurde mit der Funktion moran.test durchgeführt, da in R kein
#eigener Test für Tobit-Modelle zur Verfügung steht. Die Ergebnisse weisen
#sehr stark auf räumliche Autokorrelation hin.
#####
```

```
#####  
#####  
#####4) SAR-Modell (N=1347)#####  
#####  
#####
```

```
library(spdep)  
#Inverse-Distance min1:  
model.sar.min1 <- lagsarlm(oe1kAntLog ~  
      DNSB  
      + SI5  
      + I(MT * 10) + I((MT * 10)^2)  
      + I(MNs/10)  
      + I(MGVEproHa * 10)  
      + MKmAntLog + I(MKmAntLog^2)  
      + MAckerf1Log  
      + MLNFBetrLog  
      ,listw = W.ges.min1  
      ,data = gemeindedaten@data  
)
```

```
#Ergebniszusammenfassung:  
summary(model.sar.min1)  
bptest.sarlm(model.sar.min1)  
names(model.lag.min1)
```

```
#Berechnung der Impact-Measures  
imp.sar.min1 <- impacts(model.sar.min1, listw=W.ges.min1)  
imp.sar.min1
```

```
#Extrahieren der gefitteten Werte und der Residuen  
fw.sar.m1 <- c(model.sar.min1$fitted.values)  
summary(fw.sar.m1)  
res.sar.m1 <- c(model.sar.min1$residuals)  
summary(res.sar.m1)
```

```
#Berechnung der gefitteten Werte und der Residuen in Level-Form  
alpha.dach.sar <- 1/1347 * sum(exp(res.lag.m1))  
fw.sar.level <- alpha.dach.sar * (exp(fw.lag.m1)-1)  
summary(fw.sar.level)  
res.sar.level <- gemeindedaten$oe1kAnt - fw.sar.level  
summary(res.sar.level)
```

```
#####
#Händischer RESET-Test (Details siehe Eckey et al. 2011 - Kapitel 2: Seiten
#223ff):
#####

reset.model.sar.min1 <- lagsarlm(Oe1kAntLog ~
                                DNSB
                                + SI5
                                + I(MT * 10) + I((MT * 10)^2)
                                + I(MNs/10)
                                + I(MGVEproHa * 10)
                                + MKmAntLog + I(MKmAntLog^2)
                                + MAckerf1Log
                                + MLNFBetrLog
                                + I(fw.lag.m1^2)
                                + I(fw.lag.m1^3)
                                , w.ges.min1
                                ,data= gemeindedaten
)
summary(reset.model.sar.min1)
resettest.model.sar.min1 <- anova(model.sar.min1, reset.model.sar.min1)
#####

#####
#####4) SAR-Tobit-Modell (N=1347)#####
#####
library(spatialprobit)

#Inverse-Distance min1:
model.sartobit.min1 <- sartobit(Oe1kAntLog ~
                                DNSB
                                + SI5
                                + I(MT * 10) + I((MT * 10)^2)
                                + I(MNs/10)
                                + I(MGVEproHa * 10)
                                + MKmAntLog + I(MKmAntLog^2)
                                + MAckerf1Log
                                + MLNFBetrLog
                                ,w=Mat.w.ges.min1.sparse
                                ,data = gemeindedaten@data
                                ,showProgress=TRUE
)

summary(model.sartobit.min1.test)
names((model.sartobit.min1))
#####

#####
#Gefittete Werte händisch berechnen
#####
#Händische Berechnung der Fw und Res:
attach(gemeindedaten@data)
X
  <- cbind(as.numeric(levels(DummyGGANOE))[DummyGGANOE],
            as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld],
            as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2), I(MNs/10),
            I(MGVEproHa*10), MKmAntLog, I(MKmAntLog^2), MAckerf1Log, MLNFBetrLog)
```

```

summary(x)
dim(x)
detach(gemeindedaten@data)

y_stern <- model.sartobit.min1$y
wy <- lag.listw(w.ges.min1, y_stern)
alpha <- model.sartobit.min1$coefficients[1]
beta <- model.sartobit.min1$coefficients
beta <- beta[-length(beta)] #ausscheiden von rho
beta <- beta[-length(beta)] #ausscheiden von sigma
beta <- beta[-1] #ausscheiden vom intercept
length(beta)
beta
rho <- model.sartobit.min1.level$rho
rho

##Unzensiert:
#Formel zur händischen Berechnung der gefitteten Werte:
fw.sartobit.m1.uc <- c(rho * wy + alpha + X %*% beta)
summary(fw.sartobit.m1.uc)
#Formel zur händischen Berechnung der Residuen:
#auf Basis von y stern:
res.sartobit.m1.uc <- y_stern - fw.sartobit.m1.uc
summary(res.sartobit.m1.uc)

##Zensiert:
#Formel zur händischen Berechnung der gefitteten Werte:
fw.sartobit.m1.uc[fw.sartobit.m1.uc<=0]=0
fw.sartobit.m1.c <- fw.sartobit.m1.uc
summary(fw.sartobit.m1.c)
fw.sartobit.m1.c <- c(rho * wy + alpha + X %*% beta)
#Formel zur händischen Berechnung der Residuen:
res.sartobit.m1.c <- gemeindedaten$oelkAntLog - fw.sartobit.m1.c
summary(res.sartobit.m1.c)

#Fw und Residuen in Level-Form:
alpha.dach.sartobit <- 1/1347 * sum(exp(res.sartobit.m1.c))
fw.sartobit.c.level <- alpha.dach.sartobit * (exp(fw.sartobit.m1.c)-1)
summary(fw.sartobit.c.level)
res.sartobit.c.level <- gemeindedaten$oelkAnt - fw.sartobit.c.level
summary(res.sartobit.c.level)

#####
#Händischer RESET- und Breusch-Pagan-Test (siehe Eckey et al. 2011 -
#Kapitel 2: Seiten 223ff sowie Kapitel 2: Seiten 107ff):
#####
model.sartobit.min1.m1 <- lagsarlm(y_stern ~
                                DNSB
                                + SI5
                                + I(MT * 10) + I((MT * 10)^2)
                                + I(MNs/10)
                                + I(MGVEproHa * 10)
                                + MKmAntLog + I(MKmAntLog^2)
                                + MAckerf1Log
                                + MLNFBetrLog
                                ,listw=W.ges.min1
                                ,data = gemeindedaten@data
)

```

```

summary(model.sartobit.min1.m1)
bptest.sarlm(model.sartobit.min1.m1)

reset.model.sartobit.min1.1 <- lagsarlm(y_stern ~
      DNSB
      + SI5
      + I(MT * 10) + I((MT * 10)^2)
      + I(MNs/10)
      + I(MGVEproHa * 10)
      + MKmAntLog + I(MKmAntLog^2)
      + MAckerf1Log
      + MLNFBetrLog
      + I(fw.sartobit.m1.uc^2)
      + I(fw.sartobit.m1.uc^3)
      ,listw=W.ges.min1
      ,data = gemeindedaten@data
)
summary(reset.model.sartobit.min1.1)
resettest.model.sartobit.min1.m1 <- anova(model.sartobit.min1.m1,
      reset.model.sartobit.min1.m1)
#####

#####
#Händische Berechnung der marginal effects für das SAR-Tobit-Modell,
#beruhend auf der Formel in LeSage und Pace (2009, 38) und den in R
#implementierten Funktionen für die Marginal-Effects-Berechnung für SAR-
#Modelle (impacts.sarlm) sowie der Formel für die Berechnung des Average
#Partial Effect für Tobit Modelle aus wooldridge (2009, 591, im letzten
#Absatz. Genauere Hinweise für die Anwendung der Adaption der SAR-MEFF zu
#SAR-Tobit-MEFF finden sich außerdem in LeSage und Pace 2009, Seiten 302
#und 304.)

model.sartobit.meff <- lagsarlm(y_stern ~
      DNSB
      + SI5
      + I(MT * 10) + I((MT * 10)^2)
      + I(MNs/10)
      + I(MGVEproHa * 10)
      + MKmAntLog + I(MKmAntLog^2)
      + MAckerf1Log
      + MLNFBetrLog
      ,listw=W.ges.min1
      ,data = gemeindedaten@data
)
summary(model.sartobit.meff)
bptest.sarlm(model.sartobit.meff)

#rho wird separat abgespeichert und sigma ebenfalls, um die werte dan in
#ein "sarlm-Objekt" einfügen zu können.
rho <- model.sartobit.min1$rho
sigma <- model.sartobit.min1$sige

#Das sarlm-Objekt "model.sartobit.meff" dient als Träger für die obigen
#werte. Dazu werden die Koeffizienten und das "rho" von
#"model.sartobit.meff" einfach durch die obigen werte ersetzt. Diese
#Vorgehensweise ermöglicht es marginal effects für das sartobit Modell zu
#berechnen - da dies im Paket "spatialprobit" und der Funktion "sartobit"
#nicht implementiert ist.

```

```
koef.sartobit <- model.sartobit.min1$coefficients
koef.sartobit <- koef.sartobit[-length(koef.sartobit)] #ausscheiden von rho
koef.sartobit <- koef.sartobit[-length(koef.sartobit)] #ausscheiden von
sigma
koef.sartobit
length(koef.sartobit)
summary(model.sartobit.min1)
model.sartobit.meff$coefficients <- koef.sartobit
names(model.sartobit.meff$coefficients) <-
names(model.lag.min1$coefficients)
model.sartobit.meff$rho <- rho
summary(model.sartobit.meff)

#Die Summary zeigt, dass die Werte erfolgreich ersetzt wurden. Wichtig ist
#jedoch zu beachten, dass die restlichen Werte noch aus dem ursprünglichen
#Objekt stammen, weshalb große Vorsicht geboten ist!!!
#Da für die Berechnung der Marginal effects lediglich die beta und rho
#benötigt werden, ist diese Berechnung aber trotzdem möglich.

#Händische Extrahierung des Codes für die Berechnung:
S_inv <- invIrW(W.ges.min1, rho)
P <- matrix(koef.sartobit, ncol = 1)
n <- length(model.sartobit.meff$residuals)

lagImpacts.sar <- function (S_inv, P, n)
{
  direct <- sapply(P, function(x) sum(diag(x * S_inv))/n)
  total <- sapply(P, function(x) sum(x * S_inv)/n)
  indirect <- total - direct
  list(direct = direct, indirect = indirect, total = total)
}

test <- lagImpacts.sar(S_inv, P, n)
test
#Das Ergebnis der extrahierten Berechnung wird als "test" abgespeichert.

#Vergleich von "test" mit der für sarlm-Objekte implementierten Funktion
"impacts":
imp.sartobit.meff <- impacts(model.sartobit.meff, listw=W.ges.min1)
imp.sartobit.meff
str(imp.sartobit.meff)
direct.sartobit <- imp.sartobit.meff$direct
indirect.sartobit <- imp.sartobit.meff$indirect
total.sartobit <- imp.sartobit.meff$total

#Beide Berechnungen führen zum selben Ergebnis. Sie sind also korrekt und
#es wurde gleichzeitig festgestellt, dass für die Berechnung lediglich die
#Werte "beta (=Regressionskoeffizienten)", "n=(Anzahl der Beobachtungen"
#und SW (=SAR Generating Operator nach der Formel  $(I_n - \rho * W)^{-1}$ , welche
#in R mit der Funktion invIrM() berechnet
#werden kann) notwendig sind.

#Da es sich beim Modell um ein SAR-Tobit-Modell handelt, müssen die
#ermittelten Effekte noch berücksichtigen, dass es sich um eine Tobit-
#Spezifikation eines SAR-Modells handelt. Daher wird der average partial
#effect der ermittelten Werte berechnet (siehe Wooldridge 2009, 591. Er
#folgt der Formel:  $1/n * \text{Summe}(X \% \% \text{beta\_dach}/\text{sigma\_dach})$ ).
```



```

#Überprüfen, ob sigma noch korrekt gespeichert ist:
sigma
#Überprüfen, ob die Koeffizienten noch korrekt gespeichert sind:
koef.sartobit
names(koef.sartobit) <- names(model.lag.min1$coefficients)
#Bildung der X-Matrix ohne Intercept:
attach(gemeindedaten@data)
X <- cbind(as.numeric(levels(DummyGGANOE))[DummyGGANOE],
            as.numeric(levels(DummyGGASTmkBgld))[DummyGGASTmkBgld],
            as.numeric(levels(SI5))[SI5], I(MT*10), I((MT*10)^2), I(MNs/10),
            I(MGVEproHa*10), MKmAntLog, I(MKmAntLog^2), MackerflLog, MLNFBetrLog)
summary(X)
dim(X)
detach(gemeindedaten@data)

#Mit den bisherigen Daten den Skalierungsfaktor berechnen: Formel ist eine
#Kombination aus LeSage und Pace (2009, 294 - Gleichung Nr. 10.10) sowie
#Wooldridge (2009, 591 - APE)
summary(X)
beta <- koef.sartobit
scaling.factor.APE.sartobit <- 1/1334 * sum(pnorm(S_inv %*% I_n %*% X %*%
        beta/sigma))
scaling.factor.APE.sartobit

#Der so berechnete Skalierungsfaktor wird dann zur Funktion
#„lagImpactsExact“ hinzugefügt, indem er multipliziert wird. Die daraus
#gebildete Funktion wird „lagImpacts.sartobit“ genannt.
lagImpacts.sartobit <- function (S_inv, beta, n,
        scaling.factor.APE.sartobit)
{
  direct.sartobit <- sapply(beta[-1], function(x) sum(diag(x * S_inv *
        scaling.factor.APE.sartobit))/n)
  total.sartobit <- sapply(beta[-1], function(x) sum(x * S_inv *
        scaling.factor.APE.sartobit)/n)
  indirect.sartobit <- total.sartobit - direct.sartobit
  list(Direkt = direct.sartobit, Indirekt = indirect.sartobit, Gesamt =
        total.sartobit)
}

Marg.Eff.SARTobit <- lagImpacts.sartobit(S_inv, beta, n,
        scaling.factor.APE.sartobit)
Marg.Eff.SARTobit

Marg.Eff.SARTobit.Tab <- as.data.frame(matrix(unlist(Marg.Eff.SARTobit),
        nrow=length(koef.sartobit[-1]), ncol=3, byrow=FALSE))
Marg.Eff.SARTobit.Tab
names(Marg.Eff.SARTobit.Tab) <- c("Direkte Effekte", "Indirekte Effekte",
        "Gesamteffekte")
row.names(Marg.Eff.SARTobit.Tab) <- c(
  "g.g.A.-Gebiet Niederösterreich",
  "g.g.A.-Gebiet Steiermark und Burgenland ",
  "Spezielle Infratraktur im Umkreis von 5 km",
  "MT*10",
  "(MT*10)^2",
  "MNs/10",
  "MGVEproHa*10",
  "MKmAntLog",
  "MKmAntLog^2",

```

```

    "MAckerf1Log",
    "MLNFBetrLog"
  )

Marg.Eff.SARTobit.Tab.round <- round(Marg.Eff.SARTobit.Tab, digits=5)
Marg.Eff.SARTobit.Tab.round

#####
#Funktion zur Berechnung des AIC für das SAR-Tobit-Modell:
#####
#model.sartobit = berechnetes Modell
#left = linker Grenzwert (in diesem Fall 0)
#yHat = gefittete Werte des SAR-Tobit-Modells
#df = Freiheitsgrade (Anzahl der Parameter + Intercept)

AIC.sartobit <- function (model.sartobit, left, yHat, df)
{
  yVec <- model.sartobit$y
  obsBelow <- yVec<=left
  obsBetween <- yVec>left
  sigma <- model.sartobit$sige

  ll <- rep(NA, length(yVec))
  ll[obsBelow] <- pnorm((left - yHat[obsBelow])/sigma, log.p = TRUE)
  ll[obsBetween] <- dnorm((yVec - yHat)[obsBetween]/sigma, log = TRUE) -
    log(sigma)

  logLik.sartobit <- sum(ll[obsBelow]) + sum(ll[obsBetween])

  AIC.sartobit <- -2*logLik.sartobit + 2 * (df)
  AIC.sartobit
}

AIC.sartobit(model.sartobit=model.sartobit.min1, left=0,
             yHat=fw.sartobit.m1.uc, df=14)
#####

#####
#Diagnostikplots
#####
m <- as.data.frame(cbind(gemeindedaten$OelkAntLog, model.sartobit.min1$y,
                        fw.sartobit.m1.uc, fw.sartobit.m1.c, res.sartobit.m1.uc,
                        res.sartobit.m1.c))
names(m) <- c("y", "y*", "fw.uc", "fw.c", "res.uc", "res.c")
summary(m)
par(mfrow=c(1,2))
plot((m$fw.uc), m$res.uc, xlab="", ylab="", title(main = "Gefittete Werte
vs. Residuen", xlab="Gefittete Werte", ylab="Residuen", cex.main=2,
cex.lab=1.5))

row.names(m) <- gemeindedaten$G_Name_shp
str(m)
qqPlot(m$res.uc, id.n=5, id.cex=1.2, labels = row.names(m), xlab="Normale
Quantile", ylab = "Residuen", main = "Normalverteilungsplot der
Residuen", line="robust")
#####

```