

Vorbereitende Forschung für die zweite Evaluationsrunde Mindestlöhne

Verbesserung und Erweiterung der
Evaluationsmethoden

Mannheim, 9. Dezember 2013

Zentrum für Europäische Wirtschaftsforschung,
Mannheim

Unterauftragnehmerin: Universität Freiburg



ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Ansprechpartnerin

PD Dr. Nicole Gürtzgen (ZEW)

L 7, 1 · 68161 Mannheim

Postfach 10 34 43
68034 Mannheim

E-Mail guertzgen@zew.de

Telefon +49 621-1235-155

Telefax +49 621-1235-225

Projektteam:

Dr. Bodo Aretz (ZEW)

JProf. Dr. Melanie Arntz (ZEW)

Prof. Dr. Holger Bonin (ZEW)

Sebastian Butschek (ZEW)

Annabelle Dörr (Universität Freiburg)

Prof. Bernd Fitzenberger, PhD (Universität Freiburg, ZEW)

Terry Gregory (ZEW)

PD Dr. Nicole Gürtzgen, Projektleitung (ZEW)

Dr. Holger Stichnoth (ZEW)

Dr. Thomas Walter (ZEW)

Unter Mitarbeit von:

Bianca Blum (Universität Freiburg)

Philipp Fessler (Universität Freiburg)

Laura Hahn (Universität Freiburg)

Julius Lagodny (ZEW)

Jan Simon Ohmstedt (ZEW)

Florian Weiss (Universität Freiburg)

Jennifer Werthwein (ZEW)

Inhalt

Inhalt i

Abbildungen	iv
Tabellen	v
1 Einleitung	1
2 Systematisierung und Bewertung der ersten Evaluationsrunde.....	5
2.1 Bewertung der Gutachten zum Mindestlohn aus den Bauhaupt- und Baunebengewerbe (Branchentyp I).....	7
2.1.1 Kontrollgruppenansätze	9
2.1.1.1 Kontrollgruppen außerhalb der Branche.....	12
2.1.1.2 Kontrollgruppen innerhalb der Branche.....	19
2.1.2 Analyseebene	21
2.1.3 Funktionale Form der Zielgröße	22
2.1.4 Datengenauigkeit.....	23
2.1.4.1 Abgrenzung des betrieblichen Geltungsbereichs.....	24
2.1.4.2 Abgrenzung des persönlichen Geltungsbereichs	25
2.1.5 Zusammenfassung	27
2.2 Bewertung der Gutachten zum Mindestlohn aus arbeitsintensiven Sektoren (Branchentyp II).....	29
2.2.1 Kontrollgruppenansätze	30
2.2.2 Datengenauigkeit.....	32
2.2.2.1 Abgrenzung des betrieblichen Geltungsbereichs.....	32
2.2.2.2 Abgrenzung des persönlichen Geltungsbereichs	34
2.2.3 Zusammenfassung	35

2.3	Bewertung der Gutachten zum Mindestlohn in Sektoren mit spezifischer Wettbewerbssituation auf dem Gütermarkt (Branchentyp III)	36
2.3.1	Kontrollgruppenansätze.....	38
2.3.1.1	Kontrollgruppen außerhalb der Branche	38
2.3.1.2	Kontrollgruppen innerhalb der Branche	39
2.3.2	Analyseebene.....	41
2.3.3	Funktionale Form der Zielgröße.....	42
2.3.4	Datengenauigkeit.....	43
2.3.5	Zusammenfassung	45
2.4	Fazit.....	46
3	Alternative methodische Ansätze zur Ex-Post-Evaluation (Nicht-Strukturelle Modelle).....	51
3.1	Varianten des DvD-Ansatzes.....	51
3.1.1	DvD kombiniert mit Matching	51
3.1.2	Changes-in-Changes (CiC)	56
3.1.3	Regionale DvD-Ansätze und inkrementelles DvD	61
3.1.4	Synthetische Kontrollgruppen	65
3.2	Alternative Verfahren zum DvD-Ansatz.....	72
3.2.1	Discontinuity-Designs	73
3.2.2	Ansätze unter Verwendung von Instrumentenvariablen	80
4	Methodische Ansätze zur Ex-Ante-Evaluation (Strukturelle Modelle)	85
4.1	Strukturelle ökonomische Schätzung von Suchmodellen	88
4.1.1	Hintergrund: Suchmodelle im partiellen Gleichgewicht.....	89
4.1.2	Grenzen des Partialansatzes	90
4.1.3	Modellierung des Arbeitgeberverhaltens.....	92

4.1.4	Anpassung an die Daten	93
4.1.5	Theoretische Wirkungen eines Mindestlohns.....	94
4.1.6	Bargaining-Modelle	97
4.1.7	Empirische Evidenz	98
4.1.8	Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne.....	101
4.2	Berechenbare Allgemeine Gleichgewichtsmodelle.....	88
4.2.1	Ziele.....	105
4.2.2	Grundstruktur	106
4.2.3	Vorgehensweise und Simulationen	111
4.2.4	Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne.....	111
5	Gesamtfazit für die Erweiterung der Evaluationsmethoden.....	115
5.1	Branchentyp I (Bau und Baunebengewerbe)	117
5.2	Branchentyp II (Arbeitsintensive Sektoren)	121
5.3	Branchentyp III (Sektoren mit spezifischer Wettbewerbssituation auf dem Gütermarkt).....	123
6	Literatur	128
7	Anhang.....	138

Abbildungen

Abbildung 1: Verlauf von Y in Treatment- und synthetischer Kontrollregion ...	68
Abbildung 2: Beispiel für ein scharfes RD-Design	74
Abbildung 3: Vorgehensweise bei CGE-Modellen	114

Tabellen

Tabelle 1: Übersicht der Studien zum Branchentyp I.....	9
Tabelle 2: Placebo-Experimente in der Studie zum Bauhauptgewerbe.....	14
Tabelle 3: Übersicht der Studien zum Branchentyp II.....	30
Tabelle 4: Übersicht der Studien zum Branchentyp III.....	38
Tabelle 5: Bewertung der Ansätze nach Branchentyp	127

1 Einleitung

Die Einführung des Arbeitnehmerentsendegesetzes (AEntG) im Jahre 1996 hat die rechtliche Grundlage gelegt, tarifvertragliche Bestimmungen per Allgemeinverbindlicherklärung oder Rechtsverordnung auf alle Beschäftigten einer Branche auszuweiten. Während das Gesetz zunächst nur das Bauhaupt- und zwei Baunebengewerbe betraf, wurden bis zum Jahre 2010 insgesamt elf Branchen für jeweils unterschiedliche Geltungszeiträume in das AEntG aufgenommen. Die Auswirkungen des Mindestlohns auf Beschäftigung, Arbeitnehmerschutz und Wettbewerb in acht der elf Branchen wurden im Auftrag des BMAS und vor dem Hintergrund des § 24 AEntG durch verschiedene Forschungseinrichtungen evaluiert. Die Ergebnisse lagen im November 2011 vor.

Der im § 24 AEntG formulierte Evaluationsauftrag der Auswirkungen von Mindestlohnregelungen besteht auch über diese erste Evaluationsrunde hinaus. Für das Jahr 2014/2015 ist daher eine zweite Evaluationsrunde geplant. Diese soll die Erfahrungen der ersten Evaluationsrunde nutzen, um die Evaluationen der Mindestlohnwirkungen zu verbessern und das Evaluationsinstrumentarium gegebenenfalls zu erweitern.

Die in der ersten Evaluationsrunde untersuchten Branchen lassen sich in drei Gruppen zusammenfassen (für eine detailliertere Darstellung siehe hierzu Kapitel 2). Die erste Gruppe umfasst mit dem Bauhauptgewerbe, Malern und Lackierern, dem Dachdecker- und Elektrohandwerk Branchen des Bauhaupt- und Baunebengewerbes mit einem hohen Facharbeiteranteil. Zur zweiten Gruppe gehören die Wäschereidienstleistungen im Objektkundengeschäft sowie die Gebäudereinigung, die durch eine hohe Arbeitsintensität und einfache, standardisierte Dienstleistungen geprägt sind. Die dritte Gruppe der Branchen schließlich besteht aus der Abfallwirtschaft und der Pflegebranche – zwei Branchen, die durch eine sehr spezifische Situation auf den relevanten Gütermärkten gekennzeichnet sind. Diese Typisierung basiert zwar auf den bisher evaluierten acht Branchen, eignet sich aber prinzipiell auch dazu, die meisten der in der Zwischenzeit in das AEntG aufgenommenen Branchen (Aus-

und Weiterbildungsdienstleistungen, Gerüstbauer, Steinmetze sowie das Wach- und Sicherheitsgewerbe) oder zukünftige mögliche Branchen, wie etwa das Friseurhandwerk zu erfassen.¹

Trotz der heterogenen Ausgangslage verwenden die aus der ersten Evaluationsrunde vorliegenden Studien fast ausschließlich einen Differenz-von-Differenzen-Ansatz (DvD), um die kausalen Auswirkungen des Mindestlohns auf die zu betrachtenden Zielgrößen abzuschätzen. Der DvD-Ansatz erlaubt die Abschätzung der kausalen Wirkungen eines Mindestlohns, wenn die Veränderung in der durch den Mindestlohn betroffenen Gruppe („Treatmentgruppe“) nach dessen Einführung oder Veränderung ohne den Mindestlohn der Entwicklung in der vom Mindestlohn nicht betroffenen Kontrollgruppe entsprochen hätte und die Abweichung der Differenzen somit kausal auf den Mindestlohn zurückzuführen sind.

Um die Erfahrungen der ersten Evaluationsrunde für weitere Evaluationsprojekte zu nutzen, müssen die bestehenden Ergebnisse auf Basis des DvD-Ansatzes zunächst im Hinblick auf ihre interne und externe Validität beurteilt werden. Dafür gilt es Beurteilungskriterien zu entwickeln, die sich an den möglichen Ursachen einer Verzerrung der Ergebnisse orientieren. Die Qualität der Ergebnisse muss zudem vor dem Hintergrund der internationalen Forschungsliteratur und des dort verwendeten Spektrums an Evaluationsmethoden bewertet werden. Insbesondere ist die Frage zu klären, ob die bestehenden methodischen Möglichkeiten ausgeschöpft wurden, oder ob die jeweilige Ausgangslage der Branchenmindestlöhne alternative kausalanalytische Methoden ermöglichen würde, die die bisherigen Ansätze sinnvoll ergänzen oder diesen sogar überlegen sind.

¹ So können das Wach- und Sicherheitsgewerbe sowie das Friseurhandwerk der zweiten Kategorie zugeordnet werden, während Gerüstbauer der ersten Gruppe angehören. Die Aus- und Weiterbildungsdienstleistungen bilden durch den Dienstleistungscharakter und ein Nebeneinander von privaten und öffentlichen Bildungsträgern eine Mischform aus zweiter und dritter Kategorie.

Vor diesem Hintergrund verfolgt das vorliegende Forschungsvorhaben das Ziel, möglichst konkrete methodische Optionen aufzuzeigen und Handlungsvorschläge für die Durchführung zu machen. Der vorliegende Bericht ist hierzu wie folgt gegliedert. Kapitel 2 systematisiert zunächst die Methodik der ersten Evaluationsrunde und erarbeitet Beurteilungskriterien, die eine Einschätzung der Aussagekraft der Ergebnisse ermöglichen sollen. In Kapitel 3 und 4 sollen methodische Ansätze zur Evaluation von Mindestlöhnen erschlossen werden, die über die in der ersten Evaluationsrunde bereits eingesetzten Verfahren hinausgehen. Hierbei ist vorgesehen, sowohl Methoden, die bereits in der Mindestlohnforschung verwendet wurden, als auch solche, die bislang noch nicht im Kontext der Evaluation von Mindestlöhnen eingesetzt wurden, zu berücksichtigen. Hierbei kann es sich auch um Ansätze handeln, die aus anderen Fachgebieten als der empirischen Arbeitsmarktforschung stammen, aber potenziell einen Beitrag zur Evaluation von Branchenmindestlöhnen leisten können.

Die Erschließung potenzieller neuer Evaluationsansätze basiert auf einer umfassenden Recherche der Literatur zur empirischen Mindestlohnforschung sowie der methodischen Literatur zur Evaluation von Maßnahmen. Unterstützt wurde die Erarbeitung der Methoden zudem durch Kontakte zu Evaluationsexpertinnen und -experten verschiedener Disziplinen, die auf Basis eines kurzen Übersichtspapiers zur ersten Evaluationsrunde ihre Einschätzung zu möglichen methodischen Erweiterungen abgegeben haben.² Zudem wurden die identifizierten Methoden im Rahmen eines von der Auftragnehmerin organisierten Workshops gemeinsam mit ausgewählten externen Evaluationsexpertinnen und -experten diskutiert.³

² Das Übersichtspapier ist diesem Bericht im Anhang beigelegt.

³ An diesem Workshop haben neben dem Projektteam sowie Vertretern des BMAS als externe Berater Prof. Dr. Christina Gathmann, Prof. Dr. Wolfgang Franz, Prof. Gerard van den Berg, Ph.D. sowie Prof. Dr. Werner Wittmann teilgenommen. Wir möchten den genannten Experten und Expertinnen an dieser Stelle ausdrücklich für ihre Unterstützung danken.

Die Darstellung potenzieller alternativer Ansätze in Kapitel 3 und 4 gliedert sich in nicht-strukturelle (Kapitel 3) und strukturelle Ansätze (Kapitel 4). Nicht-strukturelle Ansätze, zu denen auch der DvD-Ansatz gehört, werden im Wesentlichen zur *Ex-Post*-Evaluation eingesetzt, indem sie die Folgen einer tatsächlich beobachteten Einführung oder Veränderung von Mindestlöhnen untersuchen, dabei aber auf eine Modellierung der generellen Wirkungszusammenhänge verzichten. Die in Kapitel 4 diskutierten strukturellen Ansätze hingegen haben das Ziel, zentrale Wirkungszusammenhänge explizit zu modellieren. Hierbei kann die abgebildete Struktur dazu verwendet werden, die Wirkungen bisher noch nicht beobachteter Rahmenbedingungen zu simulieren, was vor allem für die *Ex-Ante*-Abschätzung von Politikreformen von Interesse ist.

Die Diskussion der einzelnen Ansätze ist innerhalb der einzelnen Abschnitte in drei Schritte gegliedert. Der erste Schritt besteht zunächst in einer allgemein verständlichen Aufbereitung des jeweilig betrachteten Ansatzes. Zweitens erfolgt auf Basis der umfassenden Literaturrecherche eine Darstellung bereits bestehender Studien, die den Ansatz gegebenenfalls zur Evaluation von Mindestlöhnen verwendet haben. In einem dritten Schritt wird diskutiert, ob die betrachtete Methode sich prinzipiell für eine Folgenabschätzung der branchenbezogenen Mindestlöhne in Deutschland unter Beachtung der institutionellen Rahmenbedingungen eignet. Sofern die Anwendungsprüfung grundsätzlich negativ ausfällt, werden bestehende Hürden und Strategien zu deren möglicher Beseitigung – sofern vorhanden – explizit thematisiert.

Kapitel 5 schließt mit einem Fazit im Hinblick auf die Möglichkeiten einer verbesserten und erweiterten Evaluation der Einführung branchenspezifischer Mindestlöhne. Anders als die allgemeinen in Kapitel 3 und 4 abgegebenen Bewertungen verfolgt dieses Kapitel das Ziel, die in Kapitel 2 herausgearbeiteten unterschiedlichen Ausgangslagen der Evaluation der verschiedenen Branchen z.B. im Hinblick auf die Datengrundlage und den Evaluationszeitraum zu berücksichtigen. Hierzu werden die praktisch umsetzbaren Evaluationsmöglichkeiten zur besseren Strukturierung für die oben genannten drei Gruppen von Branchen, die durch jeweils unterschiedliche Ausgangskonstellationen gekennzeichnet sind, getrennt dargestellt.

2 Systematisierung und Bewertung der ersten Evaluationsrunde

Die Systematisierung und Bewertung der Methoden der ersten Evaluationsrunde sind wesentliche Bestandteile und die Grundlage zur Beurteilung methodischer Konzepte, die in weiteren Evaluationsprojekten genutzt werden können. Dieses Kapitel beschränkt sich dabei im Wesentlichen auf die in der ersten Evaluationsrunde durchgeführten statistischen und ökonometrischen Analysen.

Die Ausgangslage der in der ersten Evaluationsrunde betrachteten Branchen war sehr heterogen sowohl im Hinblick auf die bestehenden Marktverhältnisse als auch bezogen auf den Zeitpunkt und die konkrete Ausgestaltung der Mindestlohneinführung. Ebenso heterogen war die für Evaluationszwecke zur Verfügung stehende Datenbasis. Die Evaluation der vier Branchen des Bauhauptgewerbes und der Baunebengewerbe basiert auf umfangreichen administrativen Arbeitgeber-Arbeitnehmer-Daten. Dagegen standen für Branchen mit einer kurzen Laufzeit des Mindestlohns nur im Rahmen der Evaluation erhobene Befragungsdaten oder qualitative Forschungsdesigns zur Verfügung.

Allen Studien gemein ist die Schwierigkeit der Identifikation der tatsächlich vom Mindestlohn betroffenen Beschäftigten in den jeweiligen Branchen. Zum einen ergaben sich für einige Branchen bereits Ungenauigkeiten bei der Abgrenzung der durch die tarifvertragliche Regelung betroffenen Branche, weil die in den Sekundärdaten enthaltene Wirtschaftszweigklassifikation vielfach vom Geltungsbereich des Tarifvertrages abweicht. Zum anderen ist für die Abgrenzung der betroffenen Beschäftigten eine präzise Messung der Stundenlöhne (ohne Sonderzahlungen) notwendig. Die meisten Sekundärdaten erfassen den Stundenlohn jedoch nicht oder nur sehr unpräzise.

Trotz der heterogenen Ausgangslage wurden in der ersten Evaluationsrunde fast ausschließlich Differenzen-von-Differenzen (DvD) Ansätze zur Schätzung der kausalen Auswirkung des Mindestlohns auf vorher definierte Zielgrößen verwendet. Die wesentliche Annahme dieses Ansatzes basiert auf der parallelen Entwicklung der Ergebnisvariablen in der vom Mindestlohn betroffenen Gruppe („Treatmentgruppe“) als auch in der vom Mindestlohn nicht betroffenen Kontrollgruppe, wenn der Mindestlohn nicht eingeführt worden wäre

(Annahme paralleler Trends). Sind diese und die weiter unten diskutierten Annahmen plausibel, dann ist die Differenz in den betrachteten Zielgrößen zwischen der Treatment und Kontrollgruppe kausal auf den Mindestlohn zurückzuführen.

Die interne Validität, d.h. die eindeutige (kausale) Interpretierbarkeit der gefundenen Effekte innerhalb der untersuchten Stichprobe, basiert auf der Plausibilität der Annahme paralleler Trends. Zusätzlich muss gewährleistet sein, dass die Kontrollgruppe weder direkt, etwa aufgrund einer fehlenden Abgrenzung zwischen Treatment- und Kontrollgruppe, noch indirekt durch Skalen-, Substitutions- und Spillovereffekte durch den Mindestlohn betroffen ist.

Die oben skizzierte Heterogenität in der Ausgangslage für die Evaluation spiegelt sich in der konkreten Umsetzung der DvD-Ansätze wider. Diese unterscheiden sich je nach Zielgröße in der gewählten Analyseebene (Individual-, Betriebs- und Branchenebene), der verwendeten Kontrollgruppe, dem zugrundeliegenden Untersuchungszeitraum und der Messgenauigkeit bei der Abgrenzung der beiden Vergleichsgruppen.

Die gewählte Analyseebene sowie der zugrunde gelegte Untersuchungszeitraum beeinflussen vor allem die externe Validität, d.h. die Generalisierbarkeit der Ergebnisse. Die Wirkungen des Mindestlohns können über die Zeit im Zuge veränderter Rahmenbedingungen variieren, so dass die Ergebnisse nicht auf andere Untersuchungszeiträume übertragen werden können. Außerdem ist die Aussagekraft der gewählten Zielgröße zu prüfen. Beschäftigungswirkungen auf der Individualebene können nur bedingt mit den Wirkungen auf die Gesamtbeschäftigung gleichgesetzt werden.

Im Sinne einer übersichtlichen Darstellung der Bewertung werden die im Rahmen der ersten Evaluationsrunde untersuchten Branchen in drei Obergruppen zusammengefasst. Diese Gruppierung berücksichtigt vorhandene Strukturähnlichkeiten zwischen den einzelnen Branchen.

Zur ersten Gruppe gehören traditionelle Handwerke mit Facharbeiterdominanz, hohen Qualitätsstandards und starker Regulierung. Darunter fallen das Bauhandwerk, das Dachdeckerhandwerk, das Malerei- und Lackiererhandwerk und das Elektrohandwerk. Mindestlöhne wurden in diesen Branchen schon früh durch das Entsendegesetz eingeführt, zwischenzeitlich stark modifiziert und zum Teil in Abhängigkeit von der Branchenkonjunktur wieder ausgesetzt.

Hieraus ergibt sich eine starke Variation in den Mindestlohnregelungen, die die betreffenden Evaluationsstudien empirisch zur Identifikation der Mindestlohneffekte nutzen konnten.

Zur zweiten Gruppe gehören die Wäschereidienstleistungen im Objektkundengeschäft sowie die Gebäudereinigung. Tätigkeiten in diesen beiden Branchen sind einfache, standardisierte Dienstleistungen mit starkem körperlichem Einsatz. Charakteristisch für beide Branchen ist ein vergleichsweise hoher Anteil an weiblichen Beschäftigten. Die ersten Mindestlöhne wurden hier ab dem Jahr 2007 eingeführt.

Die Abfallwirtschaft und die Pflegebranche gehören zur dritten Gruppe. Beide Branchen sind durch ein Nebeneinander öffentlicher und privater Anbieter gekennzeichnet. Die Qualitätssicherung hat in diesen Branchen einen hohen Stellenwert. Die ersten Mindestlöhne wurden erst im Jahr 2010 eingeführt.

Im Folgenden werden die Ergebnisse der ersten Evaluationsrunde im Hinblick auf ihre interne und externe Validität anhand von Beurteilungskriterien bewertet. Die zur Bewertung herangezogenen Kriterien zielen auf das Kontrollgruppendesign (und damit die gewählte Evaluationsmethode), die gewählte Analyseebene, die funktionale Form der Zielgröße sowie die Datengenauigkeit. Neben einer Darstellung der Umsetzung der methodischen Analysen im Text geben wir in komprimierter tabellarischer Form einen Überblick über die methodischen Ansätze der jeweiligen Studien (siehe Tabelle 1, 3 und 4). Die von uns vorgenommene Beurteilung und Bewertung hinsichtlich der oben genannten Kriterien heben wir im Text kursiv hervor. In Abschnitt 2.1 diskutieren wir methodische Grundlagen, die zum Verständnis notwendig sind. Diese Erläuterungen gelten auch für die Abschnitte 2.2 und 2.3.

2.1 Bewertung der Gutachten zum Mindestlohn aus den Bauhaupt- und Baunebengewerbe (Branchentyp I)

Dem Bauhaupt- und Baunebengewerbe zugehörige Branchen bestehen aus traditionellen Handwerken mit Facharbeiterdominanz. Dazu gehören die Bauindustrie (IAB, RWI, ISG), das Dachdeckerhandwerk (ZEW), das Maler- und Lackiererhandwerk (IAW) und das Elektrohandwerk (IAW). Die mit der Evaluation beauftragten Forschungsinstitute sind in Klammern angegeben.

Ein starker Qualitätswettbewerb und ein hoher Stellenwert qualifizierter Facharbeiter kennzeichnen diese vier Branchen. Bautechnische Neuerungen und technologischer Fortschritt seitens der Materialhersteller (z.B. bei Dachsystemen) führen zu steigenden Qualifikationsanforderungen der Beschäftigten und einer erhöhten Arbeitsnachfrage nach Fachkräften. Besonders im Maler- und Lackiererhandwerk wird ein Mangel an gut qualifizierten Arbeitskräften dokumentiert. Neben der saisonalen Komponente, die in allen vier Branchen von Bedeutung ist, spielt auch die konjunkturelle Komponente im Bereich des Bauhaupt- und Baunebengewerbes eine große Rolle. Es besteht eine Abhängigkeit von staatlichen Förderprogrammen insbesondere von Fördermaßnahmen zur energetischen Gebäudesanierung, die in den letzten Jahren zu einer günstigen Auftragslage in den hier betrachteten Branchen geführt hat. Insbesondere für das Bauhauptgewerbe existieren neben makroökonomischen auch politische Konjunkturzyklen, da der Staat einer der wichtigsten Nachfrager nach Bauleistungen ist.

Eine weitere Strukturähnlichkeit dieser Branchen ist eine starke Regulierung des Arbeitsmarktes. Diese besteht aus gesetzlichen Regelungen wie dem Meisterzwang und bestehenden Rahmentarifverträgen mit Regularien zu Erschwerniszuschlägen und Maßnahmen der Beschäftigungssicherung in den Wintermonaten. Trotz eines rückläufigen Trends liegt die Tarifbindung der Betriebe zwischen 50-60%. Eine weitere Besonderheit der traditionellen Handwerke ist der sehr geringe Anteil an weiblichen Vollzeitbeschäftigten mit einem Anteil von nur ca. 3%.

Trotz der vorhandenen Strukturähnlichkeiten gibt es zahlreiche Unterschiede zwischen den Branchen. Während im Bauhauptgewerbe und im Maler- und Lackiererhandwerk der Schattenwirtschaft (Schwarzarbeit) eine große Bedeutung zugesprochen wird und eine spürbare Konkurrenz durch entsandte Arbeitnehmer besteht, sind diese Aspekte in der Dachdeckerbranche und im Elektrohandwerk von geringer Bedeutung. Auch die Wettbewerbssituation ist nicht ohne weiteres vergleichbar. Im Bauhauptgewerbe als auch im Maler- und Lackiererhandwerk werden Preise durch Ausschreibungen und Bieterverfahren festgelegt. Dabei ist der Wettbewerbsdruck in Ostdeutschland und in ländlichen Gebieten besonders hoch. In der Dachdeckerbranche ist dies nur im Segment des Neubaugeschäfts zu beobachten. Es ist davon auszugehen, dass letzteres auch für das Elektrohandwerk gilt.

Tabelle 1: Übersicht der Studien zum Branchentyp I

	Dachdeckerhandwerk (ZEW)	Bauhauptgewerbe (IAB, RWI, ISG)
Mindestlohnregelung	Einführung 1997 Mehrmalige Erneuerungen zwischen 2000 und 2010 Zeitweise Perioden ohne Mindestlohnregelungen	Mindestlohn I: Einführung 1997 Mindestlohn II: Einführung 2003, Abschaffung 2009 (Ost) Regelmäßige Erhöhungen, starke Erhöhung 1999
Daten	Qualitative Interviews mit Dachdeckerbetrieben Amtliche Statistik (Umsatzsteuerstatistik, Betriebsstatistik, Beschäftigtenstatistik, Mannheimer Unternehmenspanel (MUP) Administrative Individualdaten Daten der Lohnausgleichskasse des Daten der Bundesagentur für Arbeit (BA-Daten) CATI-Unternehmensbefragung (CATI)	Eigene Beschäftigtenbefragung Mikrozensus ELVIRA 2011 (Baustatistische Online-Datenbank) Administrative Linked-Employer-Employee Daten Bau Daten aus inländischem und ausländischem Meldeverfahren zur Urlaubskasse (SOKA Bau)
Kontrollgruppen	Installationsbranche Kontrollgruppen außerhalb der Branche Beschäftigte der Installationsbranche Betriebe der Installationsbranche Kontrollgruppen innerhalb der Branche Beschäftigte der Dachdeckerbranche, die bereits vor Einführung / Erhöhung des Mindestlohns mindestens auf diesem Niveau entlohnt wurden Dachdeckerbetriebe ohne mindestlohnbedingten Anstieg der Lohnkosten	1. baunah (Maler- und Lackiererhandwerk), 2. vorgelagert (Herstellung von Kalksandstein), 3. nachgelagert (West: Herstellung von Holzmöbel, Ost: Möbeltischlerei), 4. baufern (West: Papier und Pappeverarbeitung, Ost: Güterbeförderung Kfz) Kontrollgruppen außerhalb der Branche Beschäftigte aller Kontrollbranchen Betriebe aller Kontrollbranchen Kontrollgruppen innerhalb der Branche Beschäftigte des Bauhauptgewerbes, die bereits vor Einführung / Erhöhung des Mindestlohns mindestens auf diesem Niveau entlohnt wurden Betriebe aus dem Bauhauptgewerbe, die über Mindestlohn entlohnen
	Elektrohandwerk (IAW)	Maler- und Lackiererhandwerk (IAW)
Mindestlohnregelung	Einführung 1997 Abschaffung 2003 Wiedereinführung 2007	Einführung 2003 Erhöhung und Ausweitung des Mindestlohns 2009
Daten	Qualitative Befragung Amtliche Statistik (Erhebung im Ausbaugewerbe und bei Bauträgern, Kostenstrukturerhebung, Bauleistungspreisstatistik) IEB- und BHP-Daten Eigene Fragebogenumfrage Expertenbefragung	Qualitative Interviews Amtliche Statistik (Verdienststrukturerhebung, Unternehmensregister, Statistik der Bauleistungspreise) Mikrozensus Betriebs-Historik-Panel, IEB / Beschäftigungshistorik Eigene Befragung
Kontrollgruppen	1A. Verschiedene Branchen mit Mindestlohn, 1B. Verschiedene Branchen ohne Mindestlohn, 2. Elektrohandwerker in anderen Branchen Kontrollgruppen außerhalb der Branche Beschäftigte Kontrollbranchen 1A,B und 2 Kontrollgruppen innerhalb der Branche Beschäftigte des Elektrohandwerks, die bereits vor Mindestlohneinführung mindestens auf und maximal 15% über Mindestlohnniveau entlohnt wurden (Kontrollgruppe 3) Beschäftigte in Betrieben, die vor Einführung des Mindestlohns tarifgebunden waren (Kontrollgruppe 4)	1A. Verschiedene Branchen mit Mindestlohn, 1B. Verschiedene Branchen ohne Mindestlohn, 2. Maler und Lackierer in anderen Branchen Kontrollgruppen außerhalb der Branche Beschäftigte Kontrollbranchen 1A,B und 2 Kontrollgruppen innerhalb der Branche Beschäftigte des Maler- und Lackiererhandwerks, die bereits vor Mindestlohneinführung mindestens auf und maximal 15% über Mindestlohnniveau entlohnt wurden (Kontrollgruppe 3) Beschäftigte in Betrieben, die vor Einführung des Mindestlohns tarifgebunden waren (Kontrollgruppe 4)

2.1.1 Kontrollgruppenansätze

Die Schätzung des kausalen Effektes des Mindestlohns auf die relevanten Zielgrößen wurde in allen Studien des Bauhaupt- und Baunebengewerbes mit Hilfe eines Differenzen-von-Differenzen (DvD) Ansatzes durchgeführt. Um den kausalen Effekt einer Maßnahme/Treatment (hier: Mindestlöhne) zu identifi-

zieren, wird wie bei einem klassischen Experiment eine Gruppe von Betroffenen mit einer Gruppe von Nichtbetroffenen vor und nach Einführung der Maßnahme hinsichtlich einer vorher definierten Ergebnisgröße verglichen. Der Vorteil dieses Ansatzes besteht darin, dass Zeittrends oder strukturelle Einflüsse, die beide Vergleichsgruppen gleichermaßen betreffen, herausgerechnet werden.

Grundlegend zur Durchführung dieses Verfahrens ist eine präzise Abgrenzung der vom Mindestlohn betroffenen Beobachtungseinheiten („Treatmentgruppe“) und geeigneten nicht betroffenen Kontrolleinheiten. Gelingt die Identifikation dieser Vergleichsgruppen, werden die ersten Differenzen zwischen Treatment- und Kontrollgruppe vor und nach Einführung des Mindestlohns errechnet. Die Differenz dieser Differenzen identifiziert den kausalen Effekt der Mindestlohneinführung, wenn folgende Annahmen zutreffen: die Annahme paralleler Trends unterstellt eine Parallelität in der Entwicklung der betrachteten Ergebnisvariablen zwischen Treatment- und Kontrollgruppe, wenn das Treatment nicht stattgefunden hätte. Ist diese Annahme plausibel, entsteht erst durch das Treatment ein Unterschied in der Entwicklung der Ergebnisvariablen.

Auch wenn eine Kontrollgruppe mit hoher Plausibilität gefunden wurde, hängt ihre Eignung von Annahmen für die Analyse ab, die *per se* im strengen Sinn empirisch nicht überprüfbar sind. In diesem Sinne gibt es im Gegensatz zu einem reinen Experiment keine perfekte Kontrollgruppe. Da im DvD-Ansatz Politikmaßnahmen wie die Mindestlohneinführung als ein natürliches Experiment aufgefasst werden, kann die Annahme paralleler Trends nicht formal getestet werden. Eine Datenbasis zur kontrafaktischen Situation, d.h. einer Situation ohne Mindestlohn, existiert nicht. Deshalb ist es notwendig, bei der Auswahl der Kontrollgruppen genau zu bedenken, ob es Anzeichen für eine Verletzung dieser Annahme gibt. Werden zwei Branchen miteinander verglichen, kann eine Verletzung der Annahme aus einem makroökonomischen Schock resultieren, der die Branchen in unterschiedlicher Weise trifft. Eine Verletzung der Annahme bei Analysen auf Betriebsebene ist exemplarisch durch eine Innovation denkbar, wodurch innovationsfreudige Betriebe Beschäftigte durch Maschinen ersetzen. Der resultierende Unterschied im Beschäftigungsniveau würde im Rahmen eines DvD-Ansatzes fälschlicherweise auf die Mindestlohneinführung zurückgeführt werden.

Ein noch fundamentalerer Kritikpunkt an der Annahme eines natürlichen Experimentes ergibt sich aus der Möglichkeit der Politikendogenität der branchenspezifischen Mindestlohnregelungen in Deutschland. Es ist denkbar, dass die Tarifparteien in einer Branche gerade deshalb eine Mindestlohnregelung erwirken, weil sie für diese Branche eine bestimmte Entwicklung erwarten (antizipieren), die von der Entwicklung in den Kontrollgruppen abweicht, auch wenn in der Vergangenheit parallele Trends beobachtet wurden. Dieser Aspekt könnte zum Zeitpunkt der Einführung der ersten Mindestlohnregelung im Jahre 1997 eine wichtige Rolle spielen, da sich damals die Baukonjunktur deutlich und nachhaltig abschwächte (vor allem in Ostdeutschland) und gleichzeitig mit einer Zunahme von ausländischen Beschäftigten gerechnet wurde. Das mögliche Vorliegen von Politikendogenität stellt die Anwendung eines DvD-Ansatzes in Frage und wurde in keinem der Mindestlohngutachten ausführlich behandelt.

Die zweite zentrale Annahme des DvD-Ansatzes ist die sogenannte SUTVA-Annahme (Stable Unit Treatment Value Assumption). Der Treatmentstatus einer Beobachtungseinheit darf keinen Einfluss auf die Ergebnisvariablen der Beobachtungseinheiten der Kontrollgruppen haben. Zu Verletzungen der Annahme kommt es, wenn Spillovereffekte in Löhnen oder über Substitutions- oder Skaleneffekte in der Arbeitsnachfrage (siehe Franz 2013, Kapitel 4) auftreten.

Unter Spillovereffekten im engeren Sinn, sind direkte Lohneffekte für vom Mindestlohn nicht direkt betroffene Beschäftigte aufzufassen. Diese können zum Beispiel auftreten, wenn eine betriebliche Lohndifferenzierung aus Motivationsgründen für die Beschäftigten auch nach der Mindestlohneinführung aufrechterhalten wird. Spillovereffekte im engeren Sinn führen somit zu einer Unterschätzung von Beschäftigungsverlusten (einer Mindestlohnerhöhung). Um die bestehenden Lohndifferentiale vor einer Erosion durch den Mindestlohn zu schützen, müssen auch die Löhne oberhalb des Mindestlohnniveaus steigen. Substitutionseffekte treten auf, wenn relativ teurer gewordene Arbeitskräfte durch Beschäftigte aus der Kontrollgruppe ersetzt werden und somit die Gesamtbeschäftigung in der Kontrollgruppe steigt. Substitutionseffekte führen somit zu einer Überschätzung von Beschäftigungsverlusten. Schließlich treten Skaleneffekte auf, wenn die durch die Mindestlohnregelungen ausgelöste Erhöhung der Kosten der Produktion dazu führt, dass die pro-

duzierte Menge und damit die eingesetzte Beschäftigung zurückgehen. Der negative Skaleneffekt ist umso stärker, je weniger die Unternehmen dazu in der Lage sind, Lohnkostenerhöhungen auf die Absatzpreise zu überwälzen. Die Preisüberwälzungsspielräume sind umso geringer, je preisreagibler die Nachfrage nach den produzierten Gütern ist. Skaleneffekte führen insgesamt zu einer Unterschätzung von Beschäftigungsverlusten. Schließlich können Substitutions- und Skaleneffekte in der Beschäftigung wiederum indirekt eine Reaktion der Löhne für die betroffenen Beschäftigten durch Veränderung in den Knappheitsrelationen am Arbeitsmarkt auslösen. Insgesamt ist festzuhalten, dass die Richtung der Verzerrung der geschätzten Beschäftigungsverluste aus theoretischer Sicht offen ist (mehr dazu in Abschnitt 2.1.1.2).

Die DvD-Strategien der hier betrachteten Evaluationsstudien zu Mindestlöhnen im Bauhaupt- und Baunebengewerbe verwenden sehr ähnliche Kontrollgruppenansätze. Zum einen wird eine passende Kontrollbranche identifiziert, die nicht einem allgemeinverbindlichen Mindestlohn unterliegt. Zum anderen werden Beschäftigte und Betriebe aus der gleichen Branche identifiziert, die nicht oder nur in sehr geringem Maße durch den Mindestlohn betroffen sind. Beide Strategien haben Vor- und Nachteile, die im Folgenden diskutiert werden.

In den Studien zum Dachdeckerhandwerk und dem Bauhauptgewerbe wird der (inkrementelle) DvD-Ansatz zusätzlich in regionalen Paneldatenmodellen umgesetzt. Eine ausführliche Darstellung und Diskussion dieser regionalen Ansätze folgt in Kapitel 3.

2.1.1.1 Kontrollgruppen außerhalb der Branche

Mit der Wahl einer Kontrollgruppe aus einer anderen Branche sinkt das Risiko einer Verzerrung der geschätzten Effekte aufgrund von Spillovereffekten. Allerdings können auch bei vor- oder nachgelagerten Branchen Spillovereffekte auftreten, wenn auch in einem geringeren Umfang als innerhalb der direkt von der Mindestlohnregelung betroffenen Branche. Die Belastbarkeit der mittels eines DvD-Ansatzes gewonnenen Ergebnisse hängt maßgeblich von der Plausibilität der Annahme paralleler Trends ab. Es sollte sorgfältig geprüft und diskutiert werden, ob die wirtschaftlichen Entwicklungen in der vom Mindestlohn betroffenen Branche und der gewählten Kontrollbranche vor der Einführung des Mindestlohns ähnlich verliefen und ob davon auszugehen werden

kann, dass diese Ähnlichkeit weiter bestünde, hätte sich die Mindestlohnregelung nicht geändert. Die Auswahl einer geeigneten Kontrollbranche basiert in den hier betrachteten Studien auf einem Vergleich von konjunkturellen, institutionellen und strukturellen Rahmenbedingungen vor der Mindestlohneinführung. Im Folgenden werden die Vorgehensweisen zur Kontrollgruppenauswahl der einzelnen Studien vorgestellt.

Die Studie zu Mindestlöhnen im Bauhauptgewerbe nutzt zur Auswahl geeigneter Kontrollbranchen ein quantitatives Prüfverfahren. Ausschlaggebend für die Auswahl der Kontrollbranchen ist eine starke Ähnlichkeit hinsichtlich der Lohn- und Beschäftigungsentwicklung zwischen dem Bauhauptgewerbe und den potenziellen Vergleichsbranchen. Dabei wurden sowohl für West- als auch für Ostdeutschland eine vor- und nachgelagerte Branche sowie eine baunahe und bauferne Vergleichsbranche ausgewählt (vgl. Tabelle 1). Ob die ausgewählten Kontrollbranchen geeignet sind, den kausalen Effekt der Politikmaßnahme zu identifizieren hängt davon ab, ob die Annahme paralleler Trends für die jeweiligen Kontrollbranchen angenommen werden kann. Hinweise auf die Plausibilität dieser Annahme können in diesem Zusammenhang sogenannte Placebo-Experimente liefern.

Bei Placebo-Experimenten werden fiktive Treatments simuliert, die zu einem anderen Zeitpunkt als das echte Treatment stattfinden. Das Schätzergebnis, das unter oben vorgestellten Annahmen den kausalen Treatmenteffekt misst, wird für einen Zeitraum vor der tatsächlichen Politikmaßnahme berechnet. Ergeben sich dabei signifikante Effekte, obwohl das Treatment noch gar nicht stattgefunden hat, ist dies als Hinweis auf eine Verletzung der Annahme paralleler Trends zu bewerten. Da der geschätzte Effekt nicht auf das Treatment zurückgeführt werden kann, muss es andere Unterschiede zwischen den Vergleichsgruppen geben, die eine unterschiedliche Entwicklung in der Ergebnisgröße verursachen. Allerdings sind in diesem Zusammenhang mögliche Antizipationseffekte zu berücksichtigen, die dazu führen können, dass bereits vorab signifikante Auswirkungen des Treatments gemessen werden können. So ist es möglich, dass die Betroffenen bereits im Vorfeld auf eine bekannt gewordene Politikmaßnahme reagieren und Anpassungen in der Ergebnisgröße vornehmen.

Tabelle 2: Placebo-Experimente in der Studie zum Bauhauptgewerbe

abhängige Variable	Analyseebene	Innerhalb Branche		Vergleich mit Kontrollbranche		Effekte ab 1997
		West	Ost	West	Ost	
Lohnwachstum	Personen	gelingt	1995: <1%	1995: >1%	1995: <1%	ja
Lohnwachstum	Betriebe	1995: <1%	gelingt	-	1996: >1%	ja
Beschäftigungsniveau	Branche	-	-	Gelingt	Gelingt	nein
Arbeitsvolumen	Betriebe	gelingt	gelingt	Gelingt	gelingt	nein
Entlassungswahrscheinlichkeit	Personen	1996: >1%	gelingt	gelingt teilw.	gelingt teilw.	ja
Einstellungswahrscheinlichkeit	Betriebe	-	-	1995: =1%	gelingt	ja
befr. Beschäftigung	Betriebe	gelingt	-	-	-	nein
Betriebsgröße	Betriebe	1995: >1%	1995: >1%	-	-	ja
Umsatz	Betriebe	gelingt	gelingt	gelingt	gelingt	nein
Investitionen	Betriebe	gelingt	gelingt	gelingt	gelingt	nein
Marktaustrittsrate	Betriebe	1995: >1%	1995: >1%	scheitert überw. >1%	scheitert überw. >1%	ja
Tarifbindung	Betriebe	gelingt	-	gelingt	-	nein

Legende: Ein Placebo-Experiment gelingt, wenn keine von Null verschiedenen Effekte gefunden werden. Wenn ein signifikanter Effekt gefunden wurde, ist das Jahr angegeben und ob die quantitative Höhe des Effekts über 1%-Punkt liegt. In der letzten Spalte wird dokumentiert, ob die Wirkungsanalyse der Mindestlohneinführung im Jahr 1997 zu signifikanten Effekten führte.

Tabelle 2 gibt einen Überblick über die durchgeführten Placebo-Experimente im Bauhauptgewerbe. Die Einträge in der Tabelle geben den Zeitpunkt an, an dem ein signifikanter Effekt gemessen wurde und ob dessen quantitative Höhe über einem Prozentpunkt liegt. In der letzten Spalte der Tabelle wird angegeben, ob signifikante Effekte nach Einführung des Mindestlohns gemessen wurden.

Die Durchführung umfangreicher Placebo-Experimente für alle Ergebnisvariablen und auf allen Analyseebenen ist als sehr positiv zu bewerten. Für die Ergebnisvariable Beschäftigung scheint die Annahme paralleler Trends für alle gewählten Kontrollgruppen plausibel. Das Schätzverfahren liefert keine signifikanten Beschäftigungseffekte in West- und Ostdeutschland (vgl. IAB, RWI, ISG 2011, Tabelle 6.4). Es ist auffällig, dass die Placebo-Experimente vor allem dann gelingen (sprich parallele Trends werden nicht verworfen), wenn auch nach der Mindestlohneinführung keine signifikanten Effekte gemessen wurden. Die Experimente wurden auch für die Ergebnisvariable Lohnwachstum durchgeführt. Diese gelingen für die ausgewählten Kontrollgruppen außerhalb des

Bauhauptgewerbes nie und für Kontrollgruppen innerhalb der Branche nur für Westdeutschland. Die Ergebnisse liefern positive Effekte auf das Lohnwachstum in beiden Teilen Deutschlands für die Mindestlohneinführung 1997. Für die Einführung des Mindestlohns II und weitere Erhöhungen liefern die Untersuchungen keine robusten Ergebnisse (vgl. IAB, RWI, ISG 2011, S. 168ff).

Allgemein ist festzuhalten, dass die Auswahl geeigneter Kontrollgruppen nicht durch das Gelingen von Placebo-Experimenten motiviert sein darf, da dies auch von der Skalierung der Zielgröße (vgl. Abschnitt 2.1.3) abhängen kann. Des Weiteren steigt mit der Anzahl an durchgeführten Tests auch die Wahrscheinlichkeit dafür, dass eines der Placebo-Experimente gelingt. Die hier auftretende Gefahr eines „statistischen Overfittings“ wurde in keinem Gutachten diskutiert.

Im Rahmen der Evaluation des Mindestlohns im Dachdeckerhandwerk wurden drei potenzielle Kontrollbranchen untersucht. Diese wurden hinsichtlich einer Reihe von Kennzahlen zur Beschreibung der konjunkturellen Situation und des Wettbewerbs auf dem Arbeitsmarkt verglichen. Letztendlich zeigten sich zwischen dem Dachdeckerhandwerk und der Installationsbranche (vgl. Tabelle 1) die meisten Ähnlichkeiten. Ein quantitatives Prüfverfahren zur Kontrollgruppenauswahl wurde nicht angewandt.

Die Studie des ZEW verwendet nur eine Vergleichsbranche. Somit sind die gefundenen Effekte nur im Vergleich zur Installationsbranche zu interpretieren. Es werden keine Placebo-Experimente durchgeführt. Hier wäre interessant gewesen solche Ergebnisse für ein künstliches Treatment in den Jahren 1995 und 1996 zu errechnen. Denn wenn die Installationsbranche nicht dem kontrafaktischen Trend folgt, dann ist mit einer Verzerrung der Effekte zu rechnen.

Um die Robustheit der Ergebnisse zu überprüfen nutzt die Studie über Mindestlöhne im Dachdeckerhandwerk die Möglichkeit den DvD-Ansatz zu einem DvDvD-Ansatz zu erweitern. Dies ist möglich, wenn eine weitere Kontrollgruppe zur Verfügung steht. Eine potenzielle Kontrollgruppe sind Beschäftigte, die nicht in den persönlichen Geltungsbereich des Mindestlohns fallen. Diese Strategie bietet die Möglichkeit branchenspezifische Sondereffekte über die Zeit zu kontrollieren, die nicht auf die Mindestlohneinführung zurückzuführen sind. Die Studie zu Mindestlöhnen im Dachdecker-Handwerk nutzt den erweiterten DvD-Ansatz, um ebensolche Effekte auszuschließen. Die Ergebnisse scheinen diesbezüglich robust (vgl. ZEW 2011a, Abschnitt 7.2.1.2).

Die Studien zu Mindestlöhnen im Elektrohandwerk sowie im Maler- und Lackiererhandwerk nutzen mehrere potenzielle Kontrollgruppen (siehe Tabelle 1).⁴ Diese wurden nicht auf Basis eines quantitativen Prüfverfahrens ermittelt, sondern ihre Auswahl wurde sachlogisch begründet.

Die erste Gruppe besteht aus Beschäftigten vergleichbarer Branchen mit Mindestlohn (Hoch- und Tiefbau, Zimmerei und Holzbau sowie Fliesen- und Estrichleger), die nicht dem persönlichen Geltungsbereich angehören, d.h. die nicht Elektriker bzw. Maler- und Lackierer sind. Die zweite Gruppe identifiziert Beschäftigte vergleichbarer Branchen *ohne* Mindestlohn (Klempnerei, Bautischlerei, Parkettleger, Glaserei und Ausbautätigkeiten sowie Vermietung sonstiger Baumaschinen), die nicht dem persönlichen Geltungsbereich angehören. Als möglich dritte Gruppe wurden Beschäftigte identifiziert, die dem persönlichen Geltungsbereich angehören, d.h. die gleichen Tätigkeiten verrichten, aber in anderen Branchen beschäftigt sind.⁵

Zur Überprüfung der Ähnlichkeit von Treatment- und Kontrollgruppen wurden die Mittelwerte der Kontrollvariablen mittels einfacher Teststatistiken verglichen. Die Autoren der beiden Studien argumentieren, dass insgesamt nur wenig signifikante Unterschiede zwischen der Treatmentgruppe und den jeweiligen Kontrollgruppen bestehen. Die Unterschiede zeigen sich bei den unterschiedlichen Kontrollgruppen hinsichtlich verschiedener Charakteristiken.

Nach den IAW-Studien besteht somit das Potenzial einer echten Robustheitsanalyse. Die Autoren argumentieren, dass Effekte robust sind, wenn sie im Vergleich mit allen Kontrollgruppen festgestellt werden können. Die These der Autoren lautet, dass ein Effekt der unabhängig von der Wahl der Kontrollgrup-

⁴ Die Evaluationen der Mindestlöhne im Elektrohandwerk und im Maler- und Lackiererhandwerk wurden beide vom Institut für angewandte Wirtschaftsforschung (IAW) durchgeführt. Dies erklärt vermutlich, warum in beiden Studien dieselbe Strategie zur Identifikation der Vergleichsgruppen angewandt wurde.

⁵ Dies betrifft Elektroinstallateure, Fernmeldeinstallateure, Elektrogerätemontierer, Elektroingenieure und Techniker des Elektrofachs. Im Bereich des Maler- und Lackiererhandwerks betrifft dies Stuckateure, Gipser, Verputzer, Maler in Ausbautätigkeiten sowie Schildermaler.

pe auftritt ein echter Treatmenteffekt ist, da er nicht auf eine Variation der Kontrollvariablen zwischen den Vergleichsgruppen zurückzuführen ist.⁶

Der Robustheitstest in der durchgeführten Wirkungsanalyse in den IAW-Studien muss als gescheitert angesehen werden. In der finalen Spezifikation werden nicht alle Kontrollgruppen (vgl. Tabelle 1) verwendet. Außerdem unterscheiden sich die Ergebnisse je nach Wahl der Kontrollgruppen. Die Plausibilität der Annahme paralleler Trends wurde auch graphisch für einige Kontrollgruppen auf Basis von dargestellten Zeitverläufen der Lohn- und Beschäftigungsentwicklung analysiert.

In der Studie zum Elektrohandwerk wurden im Hinblick auf die Lohnentwicklung Vergleiche mit den Kontrollgruppen 1 und 2 durchgeführt. Ein gemeinsamer Trend scheint bis zur Einführung des Mindestlohns plausibel. Nach der Mindestlohneinführung driftet die Entwicklung auseinander. Die Analyse der zeitlichen Entwicklung der Weiterbeschäftigungswahrscheinlichkeit zeigt, dass die Annahme paralleler Trends für Kontrollgruppen 1 und 2 nicht plausibel ist (vgl. IAW 2011b, S. 176).

Eine ähnliche Analyse wurde auch für das Maler- und Lackiererhandwerk vorgestellt. Für Kontrollgruppen 1 und 2 erscheint die Annahme eines gemeinsamen Trends hinsichtlich der Stundenlöhne plausibel, bei Gruppen 3 und 4 ist die Untersuchung aufgrund der Datenstruktur nicht sinnvoll umsetzbar. Bei der Untersuchung der Weiterbeschäftigungswahrscheinlichkeit hingegen ist die Annahme nur für Kontrollgruppe 1 haltbar, die übrigen Gruppen weisen zu starke Schwankungen auf (vgl. IAW 2011a, S. 172).

Insgesamt ist die interne Validität der Ergebnisse aus beiden IAW-Studien als gering zu bewerten. Vor allem für das Elektrohandwerk scheint die Plausibilität der Annahme paralleler Trends wenig plausibel. In beiden Studien variieren die Ergebnisse je nach gewählter Kontrollgruppe. Zudem ergeben sich auch bei den durchgeführten Placebo-Experimenten signifikante Effekte (vgl. IAW 2011b, S. 198 ff., S.211 ff. und IAW 2011a, S.201 ff., S.211, S.215 ff.).

⁶ Vgl. IAW 2011a, S. 162 ff. und IAW 2011b, S. 161 ff.

Es ist festzuhalten, dass die Studie zum Bauhauptgewerbe als einzige Studie ein quantitatives Prüfverfahren zur Auswahl einer oder mehrerer geeigneter Kontrollgruppen genutzt hat. Ein in der internationalen Forschungsliteratur propagierter Kontrollgruppenansatz verwendet die vorhandene Variation in den Daten, um geeignete Kontrollgruppen zu finden. Diese als synthetischer Kontrollgruppenansatz (Abadie et al. 2010) bezeichnete Methode wird in Abschnitt 3.1.4 vorgestellt. Die Auswahl geeigneter Kontrollgruppen basiert dabei nicht nur auf sachlogischen Überlegungen, sondern die Daten identifizieren auf Basis eines geeigneten statistischen Kriteriums die ähnlichsten Kontrollgruppen und weist diesen entsprechende Gewichte zu. Das im Bauhauptgewerbe angewendete Prüfverfahren gewichtet hingegen alle Kontrollgruppen gleich.

Grundsätzlich besteht die Möglichkeit den DvD-Ansatz zu einem DvDvD-Ansatz zu erweitern, wenn eine weitere Kontrollgruppe zur Verfügung steht. Die Studie zu Mindestlöhnen im Dachdecker-Handwerk nutzt den erweiterten DvD-Ansatz, um branchenspezifische Effekte auszuschließen und die Robustheit der Ergebnisse zu überprüfen. Der Ansatz unterstellt Parallelität in den Veränderungen in den Zeittrends, eine Annahme deren Plausibilität schwieriger als für einfache DvD-Verfahren durch Placebo-Tests zu überprüfen ist. In den Studien des IAW wird der DvDvD-Ansatz erwähnt, aber es werden keine Ergebnisse dazu präsentiert.

In der Evaluation zum Bauhauptgewerbe wird zusätzlich eine Variation des DvD-Ansatzes verwendet, der die Effekte als Funktion der Eingriffsintensität modelliert, was prinzipiell sinnvoll ist. Der sogenannte inkrementelle DvD-Schätzer (Dolton et al. 2010) kontrolliert für die strukturellen Besonderheiten zwischen Betroffenen und Nichtbetroffenen, indem die Eingriffsintensität vor der Mindestlohneinführung in die Schätzung aufgenommen wird. Dieser Koeffizient wird für die tatsächliche Eingriffsintensität nach der Mindestlohneinführung (der sogenannte durchlaufende Bite des Mindestlohns, vgl. IAB, RWI, ISG 2011, S. 213 f.) verwendet, um in zeitvariabler Weise für die Selektion der von den sich im Zeitverlauf ändernden Mindestlohnregelungen Betroffenen zu kontrollieren. Der inkrementelle Effekt der tatsächlichen Eingriffsintensität nach der Mindestlohneinführung (Interaktionseffekt mit einer Jahresdummyvariablen) ist dann der geschätzte Effekt der Mindestlohnregelungen relativ zu dem Zeitraum vor der Mindestlohneinführung. Dies stellt somit einen kumulativen Effekt dar (hier ist ein Fehler in der Studie Dolton et al. 2010 zu konstatieren).

Kritisch ist hier weiter darauf hinzuweisen, dass die Eingriffsintensität im Zeitverlauf selbst von den Mindestlohnregelungen in der Vergangenheit beeinflusst wird (Endogenität der Eingriffsintensität) und deshalb die Selektionskorrektur auf Basis des vor der Mindestlohneinführung geschätzten Zusammenhangs verzerrt sein könnte. Weiterhin betreffen die kumuliert geschätzten Effekte die Wirkungen unterschiedlicher Mindestlohnregelungen für eine sich im Zeitverlauf verändernde Treatmentgruppe. Dies schließt, wie auch Dolton et al. (2010, Fußnote 16) indirekt konstatieren, eine kausale Interpretation der inkrementellen Effekte im Zeitverlauf explizit aus. Eine Modifikation (Beschränkung) des inkrementellen DvD-Schätzers für die Änderung zwischen zwei Zeitperioden alleine auf Basis der Eingriffsintensität in der ersten Periode als zeitinvariant unterstellte Selektion und einer Interaktion dieser Eingriffsintensität mit einer Jahresdummyvariable für die zweite Periode erscheint hier sinnvoller (siehe die ausführliche methodische Diskussion in Abschnitt 3.1.3).

Aufgrund der oben geäußerten Vorbehalte erscheinen die Ergebnisse des inkrementellen DvD-Schätzers ab dem zweiten Jahr nach Einführung des Mindestlohns als nicht belastbar.

2.1.1.2 Kontrollgruppen innerhalb der Branche

Eine alternative Kontrollgruppe stellen Beschäftigte oder Betriebe aus der vom Mindestlohn betroffenen Branche dar, die nicht oder nur in geringem Maße vom Mindestlohn betroffen sind. Dies können beispielsweise Personen sein, die schon vor der Mindestlohneinführung über dem Mindestlohn entlohnt wurden. Der Vorteil dieser Strategie liegt in der geringen Wahrscheinlichkeit einer Verletzung der Annahme paralleler Trends, da Treatment- und Kontrollgruppe aus der gleichen Branche stammen und somit den gleichen konjunkturellen, institutionellen und strukturellen Rahmenbedingungen unterliegen. Das Auftreten von Spillovereffekten (Substitutionseffekten, Skaleneffekten in der Arbeitsnachfrage) ist innerhalb einer Branche allerdings wahrscheinlicher. Dies kann wiederum zu einer nicht unerheblichen Verzerrung der Ergebnisse führen, deren Richtung unter Umständen angegeben werden kann. Fitzenberger (2009) argumentiert, dass für Kontrollgruppen innerhalb des gleichen Betriebes/der gleichen Branche, deren Lohn nur leicht über dem Mindestlohn liegt, der Substitutionseffekt den Skaleneffekt einer Mindestlohneinführung oder -erhöhung überwiegt und die durch einen DvD-Ansatz geschätzten negativen

Beschäftigungseffekte überschätzt sind. Dies bedeutet wiederum, dass ein DvD-Ergebnis mit nichtnegativen Beschäftigungseffekten für eine Kontrollgruppe aus einer Branche gegenüber Spillovereffekten in Form von Substitutions- und Skaleneffekten robust ist.

Im Rahmen aller Evaluationsstudien wurden Personen innerhalb der jeweiligen Branche identifiziert, die bereits vor der Mindestlohneinführung bzw. Erhöhung mindestens zum Mindestlohn entlohnt wurden. Dabei gibt es zwischen den Studien einige Unterschiede. Im Rahmen der Evaluation des Elektrohandwerks und des Maler- und Lackiererhandwerks (IAW) werden Personen als Kontrollpersonen identifiziert, die vor Einführung des Mindestlohns mindestens diesen und maximal 15% über dem Mindestlohn verdienten (vgl. Tabelle 1). Die Studien des Dachdeckerhandwerks und des Bauhauptgewerbes verwenden alle Personen als Kontrollpersonen, die mindestens mit dem Mindestlohn entlohnt wurden (vgl. Tabelle 1).

Die Vorgehensweise in den IAW-Studien ist vor dem Hintergrund vermutlicher Spillovereffekten in diesem Bereich der Lohnverteilung zu beleuchten. Die gewählte Lohngrenze scheint ad hoc festgelegt worden zu sein. Hier stellt sich die Frage, ob und in welcher Weise die Ergebnisse von der Höhe der Lohngrenze abhängen. Zudem stellt sich a fortiori das Abgrenzungsproblem, da keine Stundenlöhne beobachtet werden. In diesem eng definierten Bereich der Lohnverteilung besteht bei einer unpräzisen Beobachtung der Stundenlöhne die Gefahr einer fälschlichen Zuweisung in die Treatment- oder Kontrollgruppe. Wenn Spillovereffekte auftreten, ist die zweite zentrale Annahme des DvD-Ansatzes verletzt, da Rückwirkungen des Treatments auf die Kontrollgruppe auftreten. Dies ist beispielsweise denkbar, wenn aus Motivationsgründen die betriebliche Lohnspreizung nach Einführung des Mindestlohns aufrechterhalten wird und so auch über dem Mindestlohn liegende Löhne steigen. Auch Substitutions- und Skaleneffekte in der Arbeitsnachfrage könnten eine Rolle spielen (siehe oben).

Auf Betriebsebene werden Betriebe als Kontrollbetriebe verwendet, die über dem gültigen Mindestlohniveau entlohnen bzw. bei denen kein mindestlohnbedingter Anstieg der Lohnkosten dokumentiert wurde. Die Evaluationsstudien zu Mindestlöhnen in der Elektrobranche und im Maler- und Lackiererhandwerk identifizieren zusätzlich Beschäftigte aus Betrieben dieser Branchen,

die schon vor der Mindestlohneinführung tarifgebunden waren (vgl. Tabelle 1). Neben möglichen Substitutionseffekten besteht im Rahmen dieses Ansatzes auch das Risiko der Endogenität der Tarifbindung, d.h. Betriebe, die schon vor Einführung des Mindestlohns tarifgebunden waren, stellen im Hinblick auf die betrachteten Zielgrößen eine Selektion von Betrieben dar.

Die Analyse des Dachdeckerhandwerks kann sehr präzise Stundenlohninformationen benutzen, so dass eine Abgrenzungsproblematik zwischen den Vergleichsgruppen geringer einzuschätzen ist. In beiden Studien finden sich Hinweise für das Auftreten von Spillovereffekten. Diese werden in der Studie des ZEW sehr ausführlich besprochen. Eine Verzerrung der Ergebnisse ist auch hier möglich, wenn es zu indirekten positiven Auswirkungen auf die nichtbetroffenen Kontrollpersonen kommt.

Eine substantielle Analyse dieser Aspekte fehlt in allen Gutachten. Wenn Schätzungen der Substitutions- und Skaleneffekte in der Arbeitsnachfrage vorlägen und gleichzeitig die Spillovereffekte in den Löhnen erfasst würden, könnte die Verzerrung in den DvD-Schätzungen quantifiziert werden.

2.1.2 Analyseebene

Die hier betrachteten Evaluationen untersuchen die Effekte der Mindestlohneinführung auf Beschäftigten-, Betriebs-, und Branchenebene. Die Auswahl der Kontrollbranche erfolgt in allen Gutachten auf Branchenebene. Auf Personen- und Betriebsebene wird die Eignung der Kontrollbranchen nicht explizit untersucht. Es wird angenommen, dass sich diese auf die anderen Analyseebenen übertragen lässt. Die Plausibilität der identifizierenden Annahme (Annahme paralleler Trends) hängt davon ab, ob für eine (möglicherweise nicht beobachtete) Heterogenität zwischen den Vergleichsgruppen kontrolliert werden kann. Je ähnlicher die Vergleichsgruppen sind, desto eher ist die Annahme paralleler Trends gerechtfertigt. Während dies für ähnliche Beschäftigte einer Branche plausibel sein kann, muss dies für einen Vergleich von aggregierten Branchenwerten nicht gelten.

Zur Abgrenzung der Treatment- und Kontrollgruppen wird für die jeweilige Analyseebene die Betroffenheit der Beschäftigten bzw. Betriebe errechnet. In der Evaluationsstudie zum Dachdeckerhandwerk wird die Berechnung der Betroffenheitsmaße für jede Analyseebene ausführlich erklärt. Die Autoren do-

kumentieren eine ungleiche Verteilung der Betroffenheit auf individueller und betrieblicher Ebene (vgl. ZEW 2011a, S. 193). In dieser Studie werden identifizierende Annahmen zur Messung der kausalen Effekte des Mindestlohns auf den jeweiligen Analyseebenen detailliert diskutiert.

Grundsätzlich ist es auch möglich durch sogenannte Placebo-Experimente Hinweise über die Plausibilität der Annahme paralleler Trends auf allen Analyseebenen zu erhalten. Bei der Durchführung von Placebo-Experimenten wird so getan, als ob die zu untersuchende Politikmaßnahme (hier: Mindestlohneinführung) bereits früher stattgefunden hätte (vgl. Abschnitt 2.1.1.1).

In der Studie zum Bauhauptgewerbe wird eine besonders ausführliche Untersuchung mit Hilfe von zahlreichen Placebo-Experimenten bezüglich aller relevanten Zielgrößen auf allen drei Analyseebenen präsentiert. Die Placebo-Experimente gelingen, wenn auch sonst nur insignifikante Effekte gemessen werden. Es zeigt sich zudem, dass die Placebo Experimente in Ostdeutschland häufiger gelingen als in Westdeutschland. A priori könnte man bei der Verwendung von Kontrollgruppen innerhalb des Bauhauptgewerbes erwarten, dass die Annahme paralleler Trends eher erfüllt ist. Gleichwohl zeigen die Ergebnisse zahlreiche Verletzungen dieser Annahme.

Im Gegensatz dazu schätzen die Studien zum Elektrohandwerk sowie Maler- und Lackiererhandwerk Placebo-Experimente nur auf der Beschäftigtenebene, um die Annahme paralleler Trends zu rechtfertigen. Hierzu wurden die Ergebnisse einer fiktiven Mindestlohneinführung in den Jahren 1996, 2001 und 2005 simuliert. Einige Placebo-Experimente zeigen einen lohn erhöhenden Effekt der fiktiven Mindestlohneinführung (vgl. IAW 2011b, S. 198 und IAW 2011a, S. 193). Dies kann als Hinweis dafür gewertet werden, dass die Entwicklung für Beschäftigte in diesen Branchen positiver verläuft als in den Vergleichsbranchen und die im Bericht vorgelegten Effekte nach oben hin verzerrt sind.

2.1.3 Funktionale Form der Zielgröße

Die Skalierung der Zielgröße ist im Rahmen des DvD-Ansatzes von besonderer Bedeutung, da die Annahme paralleler Trends nicht gleichzeitig für eine lineare und eine nicht-lineare Skalierung der Zielgröße angenommen werden kann. Bei nicht-linearen Transformationen besteht die Möglichkeit, dass die Annahme paralleler Trends nicht gleichzeitig für eine Variable vor und nach der Ska-

lierung zutrifft. Somit hängt die Plausibilität der identifizierenden Annahme dieses Ansatzes davon ab, die „richtige Skalierung“ gefunden zu haben (Lechner 2011).

In allen betrachteten Berichten werden die Effekte des Mindestlohns auf den Arbeitnehmerschutz, die Beschäftigung und den Wettbewerb untersucht. Der Arbeitnehmerschutz wird durch Maße der Entlohnung abgebildet, d.h. durch die Lohnverteilung und die eventuell zusätzlich zur Entlohnung empfangenen Transferleistungen nach SGB II, die eine eventuelle existenzsichernde Wirkung des Mindestlohns abbilden sollen. Dabei wird die Entlohnung logarithmiert gemessen. Im Rahmen der Beschäftigung werden Weiterbeschäftigungswahrscheinlichkeiten allgemein und im Betrieb gemessen, der Bestand an Beschäftigten im Betrieb nach unterschiedlichen Gruppen von Beschäftigten sowie die Bewegungen am Arbeitsmarkt (Einstellungs- und Abgangsraten).

Ein möglicher Ansatz zu Umgehung der Skalierungsabhängigkeit der identifizierenden Annahmen ist die Anwendung des skalierungsunabhängigen Changes-in-Changes (CiC) Verfahrens, das von Athey und Imbens (2006) vorgeschlagen wurde (siehe auch Lechner (2011)). Dabei wird die kontrafaktische Verteilung der Ergebnisvariablen für die Treatmentgruppe nach Einführung des Mindestlohns auf Basis der Verteilung für die Kontrollgruppe zu diesem Zeitpunkt und der Differenz der Verteilungen vor Einführung des Mindestlohns geschätzt. Die Umsetzung dieses nicht-parametrischen Verfahrens ist sehr rechenaufwändig, insbesondere wenn gleichzeitig der Einfluss von weiteren beobachteten Variablen herauszurechnen ist, aber machbar (vgl. Abschnitt 3.1.2). Zusätzliche Umsetzungsschwierigkeiten treten bei diskret verteilten Ergebnisvariablen auf.

2.1.4 Datengenauigkeit

Datenbedingte Ungenauigkeiten bei der Abgrenzung der Vergleichsgruppen können zu einer Verzerrung der Schätzergebnisse gegen Null führen. In der internationalen Literatur ist dieser Effekt als „attenuation bias“ bekannt. Alle Studien aus dem Bauhaupt- und Baunebengewerbe nutzen amtliche als auch administrative Datensätze. Zusätzlich werden Unternehmensbefragungen sowie qualitative Interviews durchgeführt. Alle vier Studien können somit auf vergleichsweise gute und umfangreiche Datenquellen zurückgreifen. Trotzdem gibt es in allen Studien Ungenauigkeiten im Rahmen der Abgrenzung des be-

trieblichen Geltungsbereichs und aufgrund von fehlenden Stundenlohninformationen. Somit können Messungenauigkeiten eine weitere Fehlerquelle darstellen.

2.1.4.1 Abgrenzung des betrieblichen Geltungsbereichs

Die Studien nutzen vor allem das Betriebshistorik-Panel (BHP), um den betrieblichen Geltungsbereich nach dem Wirtschaftszweig abzugrenzen. Dabei besteht das Problem, dass die einzelnen Betriebe des BHP auf Basis einer Wirtschaftszweigsystematik erfasst werden, die keine trennscharfe Abgrenzung des Geltungsbereichs der Mindestlohnregelung erlaubt. Insbesondere besteht das Problem darin, dass die Zuordnung der Betriebe in einen Wirtschaftszweig nach dem Haupttätigkeitsprinzip erfolgt.⁷ Der Geltungsbereich des Mindestlohns wird durch den Tarifvertrag definiert. Ob ein Unternehmen darunter fällt, wird durch das Überwiegensprinzip geklärt. Nicht der Wirtschaftszweig entscheidet über die Gültigkeit des Mindestlohns, sondern die Tatsache, dass mehr als 50% der Arbeitszeit im Betrieb oder der selbständigen Betriebseinheit in dem jeweiligen vom Mindestlohn betroffenen Sektor verrichtet wird.

Somit kann ein Betrieb, der überwiegend Tätigkeiten des Elektrohandwerks ausübt, in den Wirkungsbereich des Mindestlohns fallen, ohne nach der Haupttätigkeitsdefinition ein Betrieb des Elektrohandwerks zu sein (ein solcher Betrieb würde in der Studie fälschlicherweise als nicht vom Mindestlohn betroffen angesehen). Umgekehrt kann ein Betrieb nach der Haupttätigkeitsdefinition ein Betrieb des Elektrohandwerks sein, ohne dass er nach dem Überwiegensprinzip in den Wirkungsbereich des Mindestlohns fällt (und würde somit in der Studie fälschlicherweise als vom Mindestlohn betroffen angesehen). Solche Fehlklassifikationen sind vermutlich selten, sind aber nicht auszuschließen und können prinzipiell zu einer Verzerrung der geschätzten Effekte der Mindestlohnregelungen führen (vermutlich einer Abschwächung).

⁷ Im Wesentlichen wird eine Tätigkeit zur Haupttätigkeit erklärt, die den größten Beitrag zur Wertschöpfung der beobachteten statistischen Einheit leistet, allerdings gibt es zahlreiche Sonder- und Einzelregelungen. Für weitere Details siehe Statistisches Bundesamt (2008), S. 23-37.

Eine weitere potenzielle Fehlerquelle in der Abgrenzung des Geltungsbereichs der Mindestlohnregelung ergibt sich aus der mehrmaligen Änderung der Systematik der Wirtschaftszweigklassifikation (WZK). Im Rahmen der Evaluation des Bauhauptgewerbes wurde deshalb untersucht, inwieweit Betriebe innerhalb einer WZK über die Zeit ihre zugewiesene Gruppe verlassen und durch die Übergänge zwischen den WZK die Gruppe wechseln. Problematisch ist dabei vor allem der Übergang zwischen der dreistelligen WZ73 und der fünfstelligen WZ93. Die Autoren stellen fest, dass nach der WZ73 etwa 7% der Betriebe im Bauhauptgewerbe fälschlicherweise als vom Mindestlohn betroffene Betriebe identifiziert werden.

Schließlich besteht in den BHP Daten das Problem einer uneinheitlichen Systematik der Vergabe der Betriebsnummern. Diese können sich bei einem Rechtsform- oder Inhaberwechsel eines Unternehmens verändern, ohne dass sich der Betrieb als solcher verändert hat. Fehlerrelevant sind auch sogenannte perforierte Betriebsnummernverläufe. Es gibt Betriebsnummern, die für kurze Zeiträume aus dem BHP verschwinden, dann aber wieder auftauchen, weil der Betrieb kurzzeitig unter die Erfassungsgrenze gefallen ist. Hier wird von der Fiktion ausgegangen, dass identische Betriebsnummern, die drei oder mehr Jahre aus dem Panel verschwunden sind, unterschiedlichen Betrieben zugeordnet werden können.

2.1.4.2 Abgrenzung der individuellen Betroffenheit

Stundenlohninformationen sind in allen hier betrachteten Studien von großer Bedeutung. Sie werden zum einen als Zielgröße verwendet und zum anderen sind diese Informationen maßgeblich zur Berechnung der persönlichen Betroffenheit durch den Mindestlohn, da der Mindestlohn als Stundenlohn vorgegeben ist. Die Löhne werden in allen vier Studien auf Basis der IEB-Daten (Integrierte Erwerbsbiografien) berechnet. In dieser administrativen Datenquelle liegen Informationen über Tagesentgelte vor, die im Rahmen von Meldungen der Arbeitgeber an die Sozialversicherungen dokumentiert werden. Die Tagesentgeltinformationen werden jährlich oder bei Änderungen des Beschäftigungsverhältnisses aktualisiert. Änderungen in der Entlohnung werden nicht eigens gemeldet, so dass die Tagesentgeltinformation als Durchschnittswert über den Meldezeitraum zu interpretieren ist.

In den administrativen Datenquellen wird erfasst, ob eine Stelle in Vollzeit oder in Teilzeit mit mehr oder weniger als die Hälfte der Normalarbeitszeit besetzt ist. Informationen über tatsächlich gearbeitete Stunden werden nicht erfasst, sondern auf Basis der Daten aus dem Mikrozensus geschätzt. Die durchschnittlichen Arbeitszeiten werden für Personengruppen mit bestimmten Charakteristiken bezüglich der Nationalität, dem Alter und dem Ausbildungsniveau als Durchschnittswert ermittelt. Da die Aufgliederung der WZK im Mikrozensus nur auf 3-Steller Ebene vorliegt, ist eine saubere Identifikation der Arbeitnehmer aus dem Bauhaupt- und Baunebengewerbe nicht immer möglich. Das Verfahren schätzt durchschnittliche Arbeitszeiten für Untergruppen, die nicht völlig homogen sind. Angesichts der Variation der individuellen Arbeitszeiten impliziert das Verfahren einen Messfehler in den unterstellten Stundenlöhnen, der wiederum die Abgrenzung der direkt vom Mindestlohn betroffenen Arbeitnehmer verfälscht. Ein Verzerrungspotenzial entsteht hierbei insbesondere, wenn sich die Arbeitszeit im Bauhauptgewerbe systematisch von der in anderen Wirtschaftszweigen unterscheidet. Insgesamt sehen die Autoren der Studien das Verzerrungspotenzial als gering an.

Um Hinweise auf die Plausibilität der Arbeitszeitschätzung zu gewinnen, vergleichen beispielsweise die Autoren der Studie aus dem Bauhauptgewerbe die letzten vorliegenden administrativen Daten aus dem Jahr 2009 mit den Ergebnissen der durchgeführten Beschäftigtenbefragung. Dabei stellen sie fest, dass für 93% der Befragten einer Abweichung kleiner als 10% vorliegt. Für 69% der Befragten liegt sie unter fünf Prozent. Insgesamt scheint die geschätzte Arbeitszeit also den wahren Wert zu treffen (vorbehaltlich Falschangaben)⁸.

Größere Diskrepanzen zeigen sich bezüglich der Stundenlöhne, die mittels Tagesentgeltvariablen und geschätzter Arbeitszeit errechnet werden. Hier weisen nur 45% der Befragten Abweichungen von weniger als 10% auf.⁹ Dies kann auf verschiedene Faktoren zurückzuführen sein. Zunächst ist ein Vergleich der administrativen Daten von 2009 und der Beschäftigungsbefragung

⁸ Vgl. IAB, RWI, ISG 2011, Kapitel 2.7.

⁹ Vgl. IAB, RWI, ISG 2011, S. 85 f.

2011 nur eingeschränkt möglich, da hier Entwicklungen (Lohnerhöhung, Jobwechsel) ausgeblendet werden könnten, die die Stundenlöhne beeinflusst haben. Außerdem sind zum Beispiel Zuschläge in den administrativen Daten enthalten, werden aber nicht explizit ausgewiesen, was zu einer Überschätzung des Lohnes (und bspw. einer Unterschätzung der Betroffenheit) führen kann. Da die Schätzung eine durchschnittliche Arbeitszeit liefert, werden Schwankungen in der Auftragslage oder Wettereinflüsse nicht berücksichtigt. Diese Faktoren führen zu einer unbeobachteten Variation in der Arbeitszeit, die unter Umständen auch Mehrarbeitszuschläge nach sich zieht. Je nach Wirkungsrichtung des Effekts auf die Arbeitszeit führt dies dann zu einer Über- bzw. Unterschätzung des Stundenlohnes (und damit zu einer Unterschätzung bzw. Überschätzung der Betroffenheit).

Die Autoren der Studie zu Mindestlöhnen im Dachdeckerhandwerk können Daten der Lohnausgleichkasse des Dachdeckerhandwerks (LAK) nutzen. Dieser Datensatz enthält Informationen über die tatsächlich entlohnte Arbeitszeit, so dass der individuelle Stundenlohn verlässlich berechnet werden kann. Zudem schätzen die Autoren ein verbessertes Prognosemodell der Stundenlöhne in den Daten der Bundesagentur für Arbeit auf Basis der Informationen der LAK-Daten.

Bei allen Analysen im Bereich des Bauhaupt- und Baunebengewerbes ergibt sich das Problem, dass sich der Einsatz von Entsandten und Leiharbeitskräften mit den vorliegenden Daten nicht hinreichend abbilden lässt. Da die Entsendung von Arbeitnehmern gerade im Bauhaupt- und Baunebengewerbe von Bedeutung ist, sollten die Ergebnisse vor diesem Hintergrund interpretiert werden.

2.1.5 Zusammenfassung

Alle Studien des ersten Branchentyps können auf eine solide und umfangreiche Datenbasis zurückgreifen, die eine aussagekräftige Wirkungsanalyse möglich macht. Trotzdem bestehen datenbedingte Abgrenzungsschwierigkeiten, der vom Mindestlohn betroffenen und nicht betroffenen Beobachtungseinheiten. Aktuellere administrative Datensätzen (SIAB) erlauben eine Identifikation der betroffenen Beschäftigten auf Basis der 5-stelligen WZK, so dass die Abgrenzungsprobleme des fachlichen Geltungsbereichs in zukünftigen Evaluationen geringer sein dürften.

Die interne Validität der Ergebnisse hängt außerdem maßgeblich von der Auswahl einer geeigneten Kontrollgruppe ab. Die Vorgehensweise der betrachteten vier Studien unterscheidet sich dabei erheblich. Die Evaluation über Mindestlöhne im Bauhauptgewerbe präsentiert eine sehr umfangreiche Wirkungsanalyse mit vielen Sensitivitäts- und Robustheitstests. Es ist die einzige Studie, die ein quantitatives Prüfverfahren zur Auswahl einer oder mehrerer geeigneter Kontrollgruppen genutzt hat. Kritisch einzuschätzen ist die Verwendung des inkrementellen DvD-Verfahrens.

Die Bewertung der internen Validität der weiteren Studien ist schwierig. Für das Dachdeckerhandwerk wurde nur eine Kontrollbranche verwendet. Es liegen keine Robustheitsanalysen mit anderen Kontrollgruppen oder Ergebnisse aus Placebo-Experimenten vor. Allerdings werden mögliche Verzerrungen der Ergebnisse aufgrund von Verletzungen der identifizierenden Annahmen sehr ausführlich untersucht. Die Studien des IAW nutzen viele Kontrollgruppen und führen Robustheitsanalysen durch. Hierbei zeigt sich, dass die Ergebnisse stark variieren und die Annahme paralleler Trends häufig verletzt scheint. Deshalb wären hier weitergehende Analysen (ggf. mit synthetischen Kontrollgruppen) notwendig, um die Validität der erzielten Ergebnisse zu stützen.

Leider wird die Möglichkeit des Auftretens von Spillovereffekten in den meisten Studien zum Bauhaupt- und Baunebengewerbe, die vor allem bei Verwendung von Kontrollgruppen aus derselben Branche auftreten können, nicht hinreichend im Hinblick auf mögliche Verzerrungen der Ergebnisse diskutiert. Eine Ausnahme ist die Studie zu Mindestlöhnen im Dachdeckerhandwerk, die das Auftreten von Spillover- und Substitutionseffekten sehr ausführlich behandelt. Die Abgrenzung der Vergleichsgruppen ist aufgrund der guten Datenlage für das Dachdeckerhandwerk als relativ präzise einzuschätzen.

Hinsichtlich der externen Validität der Ergebnisse muss darauf hingewiesen werden, dass die Branchen aus dem Bauhaupt- und Baunebengewerbe in vielen Dimensionen (Produktionsbedingungen, Wettbewerb, Beschäftigtenstruktur usw.) sehr speziell sind. Zudem agieren die dem Baugewerbe zugehörigen Branchen unter einer sehr speziellen konjunkturellen Lage nach der Wiedervereinigung. Somit ist die externe Validität dieser Studien beschränkt.

2.2 Bewertung der Gutachten zum Mindestlohn aus arbeitsintensiven Sektoren (Branchentyp II)

Eine Beschäftigung in arbeitsintensiven Sektoren ist durch einfache, standardisierte Tätigkeiten mit starkem körperlichem Einsatz gekennzeichnet. Dazu gehören die Wäschereidienstleistungen im Objektkundengeschäft (IAQ) sowie die Gebäudereinigung (IAQ).

Die hohe Arbeitsintensität schlägt sich in hohen Kosten der Betriebe nieder. Neben dem hohen Sachkostenaufwand der teuren Maschinen macht der Personalaufwand 54,3% der Gesamtaufwendungen der Wäschereidienstleister aus, 82,4% davon betrifft die Bruttoentgelte. Die Beschäftigten arbeiten meist in Schichten an teuren Waschstraßen. Die wichtigsten Marktsegmente sind Berufsbekleidung, Gesundheitswesen, Hotel- und Gaststättengewerbe sowie Reinigung von Maschinenputztüchern, Schmutzfangmatten oder anderen Wäschereidienstleistungen wie beispielsweise das Waschen von Handtuchrollen. In der Gebäudereinigung wird der Personalkostenanteil an den Gesamtkosten für das Jahr 2008 vom Statistischen Bundesamt auf ca. 72% beziffert

Beide Branchen haben in den letzten Jahren ein starkes Wachstum erlebt. Die Kundenzahl der Wäschereidienstleister im Objektkundengeschäft ist in diesem Zeitraum stetig gestiegen. Manche Segmente sind sehr stark konjunkturabhängig, wie beispielsweise die Reinigung von Berufsbekleidung, andere sind weniger stark konjunkturabhängig wie das Gesundheitswesen. Das starke Wachstum der Gebäudereinigungsbranche ist einem zunehmenden Trend zum Outsourcing von Reinigungstätigkeiten zuzuschreiben. Viele Betriebe diversifizieren ihr Leistungsspektrum in Richtung Gebäude- und Facility-Management.

Charakteristisch für beide Sektoren sind der hohe Anteil an weiblichen Beschäftigten sowie ein hoher Anteil an ausländischen Mitarbeitern. Auch wenn der Frauenanteil in den letzten Jahren leicht gesunken ist, sind immer noch knapp 75% der Beschäftigten weiblich. In der Wäschereibranche sind im Vergleich zur Gesamtbranche, die auch die chemische Reinigung beinhaltet, doppelt so viele ausländische Mitarbeiter (2008 waren es 12% der Beschäftigten) beschäftigt. Damit verbunden ist ein hoher Anteil an Teilzeitbeschäftigung. Rund 80% der Beschäftigten in der Gebäudereinigung arbeiten in Teilzeit und

insgesamt die Hälfte aller Beschäftigten ist in Minijobs tätig (vgl. IAQ 2011a, S. 83).

Die Wettbewerbsintensität ist in beiden Branchen als hoch einzuschätzen. Sie gilt in der Gebäudereinigung auch deswegen als außerordentlich hoch, weil bei den gewerblichen Kunden aufgrund des hohen Grades des Outsourcings nicht mehr viel Spielraum für das Neugeschäft bleibt. Marktanteile lassen sich demnach vorrangig auf Kosten der zahlreichen Wettbewerber gewinnen.

Tabelle 3: Übersicht der Studien zum Branchentyp II

	Gebäudereinigung (IAQ)	Wäschereidienstleistungen im Objektkundengeschäft (IAQ)
Mindestlohnregelung	Allgemeinverbindlichkeit von Tarifverträgen besteht schon seit 1970er Jahren Aufnahme ins AEntG im Jahr 2007 Erhöhung im Jahr 2008 Nur gewerbliche Beschäftigte sind betroffen, d.h. nicht alle Beschäftigten die Reinigungstätigkeiten durchführen	Aufnahme ins AEntG im Jahre 2009 Jährliche Erhöhungen bis April 2012 Identifikation der betroffenen Betriebe anhand der Wirtschaftszweiklassifikation (WZK) nicht möglich, Identifikation des Geltungsbereichs daher schwierig
Daten	Eigene Betriebsbefragung Expertenbefragungen Administrative Individualdaten Mikrozensus	Qualitative Interviews Amtliche Statistik Administrative Individualdaten Daten der Bundesagentur für Arbeit (BA-Daten)
Kontrollgruppen	1. Köche und Kellner, 2. ausgewählte Berufe des Wach- und Sicherheitsgewerbe, 3. Innenreiniger in anderen Branchen Kontrollgruppen außerhalb der Branche Beschäftigte in anderen Branchen Köche und Kellner Beschäftigte im Wach- und Sicherheitsgewerbe	Kontrollgruppen außerhalb der Branche keine Kausalanalyse durchgeführt Kontrollgruppen innerhalb der Branche keine Kausalanalyse durchgeführt

2.2.1 Kontrollgruppenansätze

Die verwendeten Kontrollgruppenansätze werden im Folgenden nur für die Gebäudereinigung diskutiert. Aufgrund einer schlechten Datenlage für Wäschereidienstleistungen im Objektkundengeschäft konnte in diesem Bereich keine ökonomische Wirkungsanalyse durchgeführt werden. Die Studie basiert daher ausschließlich auf Aussagen aus Expertengesprächen und Befragungen.

Ausgangspunkt für die Bildung der Treatment- und Kontrollgruppen im Rahmen der Evaluation zu Mindestlöhnen in der Gebäudereinigung sind zum einen die regional unterschiedlichen Tariflohnänderungen im April 2004 und zum anderen die Einführung der Mindestlöhne im Juli 2007 sowie die anschließende Erhöhung der tariflich vereinbarten Mindestlöhne ab März 2008.

Zur Bildung von Kontrollgruppen außerhalb der Gebäudereinigungsbranche wurden Kunden- und Beschäftigungsstrukturen verglichen. Das Wachgewerbe und die getränkegeprägte Gastronomie weisen diesbezüglich trotz bestehender Unterschiede die meiste Ähnlichkeit mit der Gebäudereinigungsbranche auf. In der Beschäftigungsstruktur gibt es vor allem bezüglich der Arbeitszeiten und des Geschlechteranteils Unterschiede zum Wachgewerbe. Bezüglich der Kundenstruktur sind sich die Branchen sehr ähnlich, wobei die getränkegeprägte Gastronomie deutliche saisonale- und klimabedingte Schwankungen aufweist.

Eine dritte Kontrollgruppe besteht aus Reinigungskräften, die in anderen Branchen beschäftigt sind und somit vergleichbare Tätigkeiten ausüben, aber keinen Mindestlohnanspruch besitzen. Die Heterogenität der Branchen, in denen Reinigungskräfte außerhalb des Gebäudereiniger-Handwerks beschäftigt sind, ist jedoch nur begrenzt kontrollierbar.

Die interne Validität der Ergebnisse ist kritisch zu bewerten. Zur Überprüfung der Plausibilität der Annahme paralleler Trends wurden zeitliche Entwicklungen der relevanten Ergebnisvariablen für die Treatment- und Kontrollgruppe illustriert (vgl. IAQ 2011a, S. 250 ff.). Diese branchenspezifischen Entwicklungspfade stellen die Gültigkeit der identifizierenden Annahme infrage und sprechen somit gegen eine Eignung der Kontrollgruppen. Insbesondere für die Beschäftigungsentwicklung scheint diese wichtige Annahme wenig plausibel. Es wurden keine Placebo-Experimente zur weiteren Plausibilitätsprüfung durchgeführt.

Die geschätzten Effekte könnten auch als Hinweis auf eine Endogenität des Treatments (zur Politikendogenität siehe oben) bewertet werden. Damit ist gemeint, dass die Einführung der Mindestlöhne und die spätere Erhöhung im Hinblick auf zukünftige Beschäftigungseffekte von den Tarifparteien ausgehandelt wurden. Dieses Problem stellt u.U. den gewählten Evaluationsansatz in Frage und hätte weiter untersucht werden sollen.

Die Studie zu Mindestlöhnen in der Gebäudereinigungsbranche fokussiert bei der Anwendung des DvD-Ansatzes nur auf die Regionen- und Beschäftigtenebene. Als Ergebnisvariable wird einzig die Beschäftigungswahrscheinlichkeit betrachtet. Somit wird im Folgenden ausschließlich die Datenbasis diskutiert.

Eine Darstellung und Diskussion der Analyseebenen und der funktionalen Form der Zielgröße ist hinfällig.

2.2.2 Datengenauigkeit

Im Vergleich zu der Datengrundlage der vier Studien aus dem Bauhaupt- und Baunebengewerbe gibt es im Bereich der arbeitsintensiven Sektoren grundlegende Probleme der Datenverfügbarkeit. Dies liegt zum einen an der relativ kurzen Zeitperiode seit Einführung der Mindestlöhne und dem Zeitpunkt der Evaluation. Zum anderen existieren im Wäschereigewerbe aufgrund der Verfügbarkeit der Informationen zu der Wirtschaftszweigklassifikation erhebliche Abgrenzungsprobleme der überhaupt vom Mindestlohn betroffenen Betriebe und Personen.

2.2.2.1 Abgrenzung des betrieblichen Geltungsbereichs

Aufgrund der unzureichenden Datenlage sowie der Zuordnungs- und Identifikationsprobleme der vom Mindestlohn betroffenen Betriebe wurde im Rahmen der Evaluation der Mindestlöhne im Bereich der Wäschereidienstleistungen keine Wirkungsanalyse vorgenommen. Dennoch wurden 327 Interviews durchgeführt, 181 davon mit Betrieben, die nach eigenen Angaben unter die Mindestlohnregelung fallen würden. Die Stichprobe für die Interviews wurde aus zwei Quellen erhoben. Zum einen wurde eine große Stichprobe aus der WZ 2008 für Wäscherei und chemische Reinigung (WZ 96.01.0) entnommen und gefragt, ob es sich um einen Wäschereibetrieb handelt. War dies nicht der Fall, so wurde die Beobachtung aus der Stichprobe gelöscht. Zusätzlich nahmen die Autoren der Evaluation Betriebe in die Stichprobe auf, die von Experten aus der qualitativen Analyse tendenziell als mindestlohnpflichtig bewertet wurden, obwohl sie nicht dem Wirtschaftszweig zugeordnet waren.

Die Probleme, die sich für die Autoren ergeben haben, waren vor allem die Zuordnung der Betriebe zur Mindestlohnpflicht. Die Mindestlohnregelungen betreffen Betriebe und selbständige Betriebsabteilungen, die gewerbsmäßig überwiegend Textilien für gewerbliche Kunden sowie öffentlich-rechtliche oder kirchliche Einrichtungen waschen, unabhängig davon, ob die Wäsche im Eigentum der Wäscherei oder des Kunden steht. Der Umsatzanteil mit gewerblichen Kunden sowie öffentlich-rechtlichen oder kirchlichen Einrichtungen muss größer als 80% sein. Generell ausgenommen von der Mindestlohn-

pflicht sind Werkstätten für behinderte Menschen. Die Mindestlöhne gelten mit Ausnahme von Auszubildenden für alle Beschäftigten.

Es konnte im Rahmen der Evaluation kein eindeutiges und stichhaltiges Kriterium gefunden werden, das einen Betrieb eindeutig als mindestlohnpflichtig identifiziert. Dennoch wurden zwei Kriterien ermittelt, die zumindest einen Anhaltspunkt für die Mindestlohnpflicht eines Betriebes geben konnten. Zum einen gaben über 90% der Betriebe, die sich selbst für nicht mindestlohnpflichtig hielten, eine tägliche Waschleistung von unter einer Tonne an. Die zweite Kenngröße für die mögliche Mindestlohnpflicht betrifft die Betriebsgröße. Unter den mindestlohnpflichtigen Betrieben dominieren Betriebe mit einer Beschäftigtenzahl von über 20 Mitarbeitern. Ab einer Betriebsgröße von über 100 Mitarbeitern hielten sich alle Betriebe in der Stichprobe für mindestlohnpflichtig.

Leider sind diese beiden Kriterien nicht stichhaltig genug, um sie als Indikator für die Mindestlohnpflicht einer Wäscherei zu nutzen. Es wurden im Rahmen der Evaluation Unstimmigkeiten in den Angaben über die Mindestlohnpflicht festgestellt, so dass sich Betriebe mit einem Umsatzanteil im Objektkundengeschäft unter 80% für mindestlohnpflichtig hielten, obwohl sie es im Rahmen der gesetzlichen Regelung nicht gewesen wären. Andere Betriebe, die laut ihrem Umsatzanteil eindeutig in die Mindestlohnregelung fallen würden, gaben an nicht mindestlohnpflichtig zu sein.

Die Datenlage für eine Evaluation der Mindestlohnwirkungen im Wäschereigewerbe ist sehr schlecht. Trotzdem ist eine kausale Wirkungsanalyse im Rahmen eines DvD-Ansatzes mit unvollständiger Betroffenheit denkbar. Der Betroffenheitsgrad von Betrieben in den Daten der Bundesagentur für Arbeit hätte auf Basis der beobachteten Variablen und Vergangenheitsdaten geschätzt werden können. In Verbindung mit dem in Abschnitt 2.1.1.1 erwähnten synthetischen Kontrollgruppenansatz hätte es die Möglichkeit gegeben, Ergebnisse aus einer kausalen Wirkungsanalyse zu präsentieren. Allerdings wäre dies vermutlich nicht in der Kürze der zur Verfügung stehenden Zeit machbar gewesen, da erst nach einer detaillierten Analyse der Eignung potenzieller Kontrollbranchen die notwendigen Befragungen für den Zeitraum nach Einführung des Mindestlohns für die geeigneten Kontrollbranchen hätten geplant werden können. Gerade im Hinblick auf das mögliche Problem geringer Fallzahlen in den

Kontrollbranchen (in den Befragungen) für die kleine zu untersuchende Branche Wäschereigewerbe weist der synthetische Kontrollgruppenansatz Vorteile auf (Abadie et al. 2010).

In der Evaluation der Gebäudereinigung findet das Betriebshistorik-Panel keine Anwendung, da eine Identifikation der Gebäudereinigungsbetriebe nicht möglich war. So musste berücksichtigt werden, dass in dem relevanten Wirtschaftszweig auch Betriebe erfasst sind, die nicht unter die Mindestlohnregelungen fallen oder für die eine Mindestlohnregelung nicht gilt, weil eigene Tarifstrukturen existieren. Dies ist beispielsweise bei Betrieben der Schornsteinreinigung relevant. Eine Wirkungsanalyse auf Betriebsebene wurde deshalb nicht durchgeführt.

2.2.2.2 Abgrenzung des persönlichen Geltungsbereichs und der individuellen Betroffenheit

Bezüglich der Stundenlohninformationen besteht das gleiche Problem wie in den Studien zum Bauhaupt- und Baunebengewerbe beschrieben. Deshalb spielt auch hier der Mikrozensus eine große Rolle, da er (im Gegensatz zu den Daten der Bundesagentur für Arbeit wie SIAB) Informationen zur normalerweise geleisteten und tatsächlichen Arbeitszeit der Beschäftigten enthält (bezogen auf die Woche vor der Befragung). Aufgrund der hohen Relevanz von Teilzeitarbeit in den arbeitsintensiven Sektoren ist diese Information besonders wichtig. Auf Basis der Informationen des Mikrozensus konnte die Entwicklung des Arbeitsvolumens gemessen werden.

Zusätzlich waren für die Gebäudereinigung administrative Daten über das Forschungsdatenzentrum der Bundesagentur für Arbeit verfügbar. Für die Evaluation äußerst relevant war die Stichprobe der integrierten Arbeitsmarktbiographien (SIAB). Ein großer Vorteil der SIAB ist die Angabe der Wirtschaftszweigklassifikation auf 5-Steller-Ebene. Somit konnten Beschäftigte im Bereich der Schornsteinreinigung aus der Evaluation ausgeschlossen werden, da für diese Personen die Mindestlohnregelungen der Gebäudereinigung nicht gelten.

Weil sich die Autoren für die Evaluation auf Reinigungsberufe beschränken, können einige Beschäftigte unberücksichtigt bleiben, da sie zwar unter die Mindestlohnregelung fallen, statistisch aber nicht erfasst sind. Jedoch sehen

die Autoren den großen Vorteil ihrer Vorgehensweise darin, dass wohl kaum Beschäftigte in ihre Evaluation eingehen, die nicht von der Mindestlohnregelung betroffen sind.

2.2.3 Zusammenfassung

Die Evaluationen zu Mindestlöhnen in arbeitsintensiven Branchen leiden an einer schlechten Datengrundlage und damit verbundenen Abgrenzungsschwierigkeiten der Vergleichsgruppen. In beiden Branchen lag die Mindestlohneinführung zum Zeitpunkt der Evaluation nicht weit zurück. Dies führte insbesondere für die Studie zum Mindestlohn im Wäschereigewerbe zu Schwierigkeiten.

Zum Zeitpunkt der Evaluation herrschte viel Unsicherheit auf Seiten der Wäschereibetriebe über die Mindestlohnpflicht. Hinzu kam die unzureichende Datenlage, die theoretisch mindestlohnpflichtige Betriebe nicht eindeutig identifizieren konnte. Selbst eine eigene Befragung konnte keinen ausreichenden Datensatz liefern, um die Wirkungen des Mindestlohns in einer Kausalanalyse zu prüfen. Eine kausale Wirkungsanalyse, die die unvollständige Betroffenheit von Betrieben als Treatment verwendet, wäre in Kombination mit modernen Kontrollgruppenansätzen eine Alternative gewesen.

Die kausale Wirkungsanalyse zum Mindestlohn in der Gebäudereinigung leidet unter methodischen Schwächen. Es wird nicht ausreichend begründet, warum die gewählten Kontrollgruppen die erforderlichen Annahmen des Schätzmodells erfüllen. Gerade im Hinblick auf die großen Unterschiede in der Beschäftigungsstruktur wäre eine detaillierte Untersuchung mit Placebo-Experimenten sinnvoll gewesen. Eine interne Validität der Ergebnisse ist somit nicht gegeben.

Beide Branchen sind zudem sehr speziell und nicht repräsentativ für andere Sektoren. Selbst wenn die Ergebnisse für die Gebäudereinigung intern valide wären, könnten diese nicht auf andere Branchen oder gar auf die Gesamtwirtschaft übertragen werden. Eine externe Validität der empirischen Ergebnisse erscheint nicht gegeben – insbesondere im Hinblick auf die speziellen theoretischen Umstände, unter denen die Ergebnisse interpretierbar sind.

2.3 Bewertung der Gutachten zum Mindestlohn in Sektoren mit spezifischer Wettbewerbssituation auf dem Gütermarkt (Branchentyp III)

Der dritte Branchentyp zeichnet sich durch ein Nebeneinander von privaten und öffentlichen Anbietern aus. Diesem Typ zugehörige Sektoren sind die Abfallwirtschaft (ZEW) und die Pflegebranche (IAW).

Die spezifische Wettbewerbssituation führt zu einer starken Heterogenität in diesen Branchen. In der Pflegebranche zeigt sich die Heterogenität im Bereich der Einrichtungen, die von privaten, öffentlich-rechtlichen und freigemeinnützigen Trägern geführt werden. Vor allem ambulante Einrichtungen befinden sich häufig in privater Trägerschaft, hingegen sind stationäre Einrichtungen meist in frei-gemeinnütziger Hand. Auffällig hierbei ist, dass öffentliche Träger meist nach tariflichen Vereinbarungen entlohnen, die oberhalb des Mindestlohns liegen. Im Bereich der Abfallwirtschaft ist eine Heterogenität auf den existierenden Teilmärkten zu beobachten. Die wichtigsten vier Teilmärkte der Abfallwirtschaft sind Sammlung von Abfällen, Abfallbehandlung und -beseitigung, Rückgewinnung (Recycling) und sonstige Entsorgung (u.a. Reinigung öffentlicher Verkehrsflächen wie Winterdienste). Sie unterscheiden sich vor allem im Hinblick auf Marktzutrittsbarrieren und die Qualifikationsanforderungen an die Beschäftigten.

Eine weitere Gemeinsamkeit beider Branchen ist eine vergleichsweise konstante Nachfrage. Im Bereich der Pflege besteht keine Abhängigkeit von der aktuellen Wirtschaftslage. Die Nachfrage nach Pflege folgt dem demographischen Wandel. Insbesondere für die Berufsgruppe der Pflegefachkräfte ist ein Fachkräftemangel zu verzeichnen, der zu höheren Löhnen für diese Berufsgruppe führt. Eine konstante Nachfrage ist auch charakteristisch für die Abfallwirtschaft. Einige Teilbranchen sind konjunkturunabhängig (z.B. die Hausmüllentsorgung), in anderen hängt die Nachfrage stark von den schwankenden Preisen für Sekundärstoffe wie Altmetall, Altpapier und Kunststoffe ab (z.B. Recycling). Dennoch resultiert eine Konstanz der Nachfrage aus der gesetzlichen Entsorgungsverpflichtung von Abfällen.

Nach dem Kreislaufwirtschafts- und Abfallgesetz obliegt die Hausmüllentsorgung öffentlich-rechtlichen Entsorgungsträgern. Die Hausmüllentsorgung kann entweder über öffentliche Ausschreibungen an beauftragte Dritte (d.h. auch

an private Anbieter) oder über kommunale Eigenbetriebe durchgeführt werden. Öffentliche Ausschreibungen werden häufig über die niedrigsten Angebotspreise gewonnen. Dabei schließen sich häufig größere Unternehmen zu Gemeinschaftsunternehmen und Bietergemeinschaften zusammen. Während insbesondere nicht tarifgebundene Entsorgungsunternehmen niedrige Preise bieten können, sind tarifgebundene private und kommunale Entsorgungsunternehmen mit erheblichen Wettbewerbsnachteilen konfrontiert. Im Bereich der Wiederverwertung von Sekundärstoffen sind aufgrund der hohen, wirtschaftlichen Renditen steigende Zahlen von Marktzutritten zu beobachten.

Grundsätzlich bestehen in der Abfallwirtschaft branchenintern wenige Substitutionsmöglichkeiten, weil der Mindestlohn alle Arbeitnehmer betrifft, die Branche in weiten Teilen durch eine hohe Arbeitsintensität geprägt ist und wenige Effizienzgewinne durch zusätzlichen Technikeinsatz zulässt. Auch der Pflegesektor ist sehr personalintensiv, und es bestehen nur beschränkte Substitutionsmöglichkeiten durch technologischen Fortschritt oder ungelernete Arbeitskräfte (z.B. Hauswirtschaftskräfte). Der Gesetzgeber schreibt eine Fachkraftquote von 50% vor, d.h. vor Einführung des Mindestlohns war der Einsatz von ungelerten Pflegehilfskräften begrenzt. Denkbar ist jedoch, dass der Mindestlohn zu einer Substitution von relativ teurer gewordenen ungelerten Pflegehilfskräften durch relativ billiger gewordene Fachkräfte führen kann.

Bezüglich der Beschäftigtenstruktur zeigen sich Unterschiede in den Geschlechteranteilen. Der Frauenanteil ist im Gegensatz zur Abfallwirtschaft mit ca. 80-90% im Pflegebereich sehr hoch. Laut der Pflegestatistik 2009 beträgt der Frauenanteil zwischen 85-90% (vgl. IAW 2011c, S. 77). Ein weiterer Unterschied besteht in der Möglichkeit der Kostenüberwälzung auf die Verbraucher. In der Abfallwirtschaft erlaubt das im öffentlichen Gebührenrecht verankerte Kostendeckungsprinzip im Bereich der Hausmüllentsorgung einen Kostenanstieg vollständig auf die Abfallverursacher zu überwälzen. Im Pflegebereich können gestiegene Kosten nicht einfach auf die Preise übertragen werden, da das Budget der Einrichtungen grundsätzlich Verhandlungen mit den Krankenkassen voraussetzt und sich über das Standard-Pflegesatz-Modell bildet.

Tabelle 4: Übersicht der Studien zum Branchentyp III

	Pflegebranche (IAW)	Abfallwirtschaft (ZEW)
Mindestlohnregelung	Einführung des Mindestlohns im August 2010 Erhöhung in 0,25-Euroschritten im Januar 2012 und Juli 2013	Aufnahme ins AEntG Anfang 2010 Erhöhungen Ende 2010 und 2011
Daten	Qualitative Interviews Amtliche Statistik (Pflegestatistik, Betriebs-Historik-Panel), Integrierte Erwerbsbiographien, Unternehmensregister, Verdienststrukturerhebung) Mikrozensus Befragungsdaten	Qualitative Interviews Amtliche Statistik (Umsatzsteuerstatistik, Strukturerhebung im Dienstleistungsbereich und prod. Gewerbe, Statistik Beschäftigten, Mannheimer Unternehmenspanel (MUP) CATI-Unternehmensbefragung
Kontrollgruppen	Kontrollgruppen außerhalb der Branche keine verwendet Kontrollgruppen innerhalb der Branche Einrichtungen, deren Beschäftigte nach kollektivrechtlichen Verträgen in Höhe des Mindestlohns oder knapp oberhalb des Mindestlohns entlohnt werden	Kontrollgruppen außerhalb der Branche Betriebe des Garten- und Landschaftsbaus (nur als Kontrollgruppe für Abfallbeseitigung verwendet) Kontrollgruppen innerhalb der Branche Betriebe der Abfallwirtschaft, die keine Lohnerhöhung bei den Mitarbeitern in Folge des Mindestlohns vornehmen mussten

2.3.1 Kontrollgruppenansätze

Im Folgenden werden auch für den Branchentyp III die verwendeten Kontrollgruppenansätze vorgestellt. Kontrollgruppen aus anderen Vergleichsbranchen wurden nur in der Studie zur Abfallwirtschaft vom ZEW verwendet.

2.3.1.1 Kontrollgruppen außerhalb der Branche

Als Kontrollgruppe außerhalb der Branche wurde der Garten- und Landschaftsbau (Gala-Bau) gewählt, weil diese Branche, wie die Abfallwirtschaft auch, durch Multiproduktunternehmen gekennzeichnet ist, wie beispielsweise Dienstleistungen in der Objektbetreuung aber auch Hausmeistertätigkeiten oder Ähnliches. Auch in dieser Branche gibt es eine spezifische Wettbewerbssituation durch das Nebeneinander von öffentlichen und privaten Anbietern sowie geringe Qualifikationsanforderungen an die Beschäftigten. Ausschreibungen werden ebenfalls durch Niedrigpreisangebote gewonnen, was auch den Konkurrenzdruck erhöht. Die öffentlichen Anbieter haben gegenüber den privaten Anbietern Wettbewerbsvorteile, da sie mit unterschiedlichen Umsatzsteuersätzen belastet werden.

In einem umfassenden Branchenbild wurde die Annahme paralleler Trends zwischen Abfallwirtschaft und Gala-Bau untersucht. Mögliche Verletzungen der Annahme zeigen sich vor allem in der Sektorengröße, der Umsatzentwicklung sowie der Unternehmensanzahl und deren Größe. Der Gala-Bau zeichnet sich vor allem durch viele Kleinbetriebe aus, die zudem wesentlich häufiger tarifgebunden sind als in der Abfallwirtschaft. Aus diesen Unterschieden her-

aus wurde die Kontrollgruppe in der Evaluation ausschließlich für den Teilbereich der Abfallbeseitigung verwendet.

Die Wirkungsanalyse betrachtet den Effekt der Mindestlohneinführung auf Betriebe im Teilbereich der Abfallbeseitigung im Vergleich zu Betrieben aus dem Landschafts- und Gartenbau. Die externe Validität dieses Ergebnisses ist sehr beschränkt, da es nur für den Teilbereich der Abfallbeseitigung gilt. Gerade in diesem Teilbereich der Abfallwirtschaft ist die Nachfrage vergleichsweise konstant und die Möglichkeiten Beschäftigung abzubauen oder diese zu ersetzen sind gering. Die Verwendung der weiteren Teilmärkte der Abfallwirtschaft wäre eventuell möglich gewesen, wenn die Eignung der Kontrollgruppe als Vergleichsbranche für diese Teilmärkte durch Placebo-Tests überprüft worden wäre. Außerdem hätte das Analysepotenzial unter Anwendung des synthetischen Kontrollgruppendesigns ausgenutzt werden können. Allerdings wäre dies vermutlich nicht in der Kürze der Zeit für das Gutachten zu schaffen gewesen, da dann für alternative Branchen zunächst eine detaillierte Analyse mit Placebo-Tests auf Basis administrativer Daten vor Einführung des Mindestlohns im Hinblick auf die Eignung der Kontrollbranchen hätte durchgeführt werden müssen. Erst nach Vorliegen dieser Ergebnisse, hätten die notwendigen Befragungen für den Zeitraum nach Einführung des Mindestlohns für die geeigneten Kontrollbranchen geplant werden können.

2.3.1.2 Kontrollgruppen innerhalb der Branche

In beiden Studien werden Kontrollgruppen innerhalb der untersuchten Branche definiert. Dies hat den Vorteil, dass die Annahme paralleler Trends plausibler ist. Allerdings besteht hier wie in den gleichgearteten Ansätzen für die in den Abschnitten 2.1 und 2.2 behandelten Branchen die Möglichkeit von direkten Substitutionseffekten in Löhnen oder von Substitutions- und Skaleneffekten (siehe hierzu die Ausführungen im Abschnitt 2.1.1). Auf diese Kritikpunkte wird hier nur noch punktuell eingegangen.

Die IAW-Studie für die Pflege wählt die Gruppe der tarifgebundenen Einrichtungen als Kontrollgruppe, da diese nicht vom Mindestlohn betroffen sind, aber dieselbe Entwicklung zu erwarten hatten wie die vom Mindestlohn betroffene Gruppe. Substitutionseffekte im Hinblick auf andere Branchen sind vermutlich vernachlässigbar, da die Pflegebranche branchenspezifische Tätigkeiten erfordert. Substitutionseffekte zwischen Treatment- und Kontrollgrup-

pe sind nicht zu erwarten, da die Tarifbindung eine Eigenschaft der Einrichtung ist und nicht der Arbeitnehmer. Die Arbeitnehmer aus der Kontrollgruppe (also aus Betrieben mit Tarifbindung) sind somit nicht qualifizierter oder für einen Arbeitgeber interessanter als jene in der Treatmentgruppe. Allerdings besteht die Möglichkeit der Substitution durch die Nachfrage nach Pflegeleistungen in der Kontrollgruppe, da deren Lohnkosten nicht ansteigen.

Ein Substitutionseffekt, der in der Evaluation ausführlich diskutiert, aber empirisch nicht untersucht werden konnte, betraf die Substitution mit ausländischen Arbeitern oder Scheinselbstständigen. Beide Effekte wurden mit sinnvoller Begründung ausgeschlossen. Die hohen Anforderungen an eine ausführliche Dokumentation der pflegerischen Tätigkeiten schließt die Substitution mit ausländischen Fachkräften größtenteils aus, da Sprachbarrieren bestehen. Das Problem der Scheinselbstständigkeit wurde in der Diskussion als vernachlässigbar angesehen, weil das hohe Risiko der Aufdeckung durch die zuständigen Behörden die Quote der Scheinselbstständigen marginal werden lässt.

Skalen- und Substitutionseffekte, also Nachfrageänderungen aufgrund von Änderungen der Lohnrelationen bzw. der Lohnkosten, sind wegen der Sektorenspezifika in der Pflegebranche nur in geringem Ausmaß zu erwarten. Die Arbeitsnachfrage ist vergleichsweise unelastisch. Die Arbeitskräfte werden benötigt und können nicht durch Maschinen ersetzt werden. Eine Substitution durch Hauswirtschaftskräfte ist nur bedingt möglich. Auch ist im Hinblick auf die Preisbildung über das Standard-Pflegesatz-Modell zu beachten, dass Preisadjustierungen aufgrund von Kostenänderungen Verhandlungen mit den Pflegekassen voraussetzen (die IAW-Studie spricht von zähen und langen Verhandlungen).

Placebo-Experimente zur Überprüfung der identifizierenden Annahmen konnten nicht mit demselben Datensatz durchgeführt werden, da die Panelbeobachtungen nur die zwei Perioden umfassen, die für die Wirkungsanalyse verwendet wurden. Die Autoren nutzten deshalb Lohninformationen aus administrativen Individualdaten. Somit wurden die Placebo-Experimente auf Beschäftigtenebene durchgeführt und sind nicht mit den Ergebnissen der Wirkungsanalyse auf Betriebsebene vergleichbar. Die grundsätzliche Durchführung solcher Robustheitsanalysen ist deutlich zu befürworten. In diesem Fall

tragen sie jedoch nur wenig zur Bewertung der internen Validität der Ergebnisse bei.

In der Evaluation nicht hinreichend diskutiert wurde die mögliche Substitution zwischen Pflegehilfs- und Pflegefachkräften. Sie wurde als vernachlässigbar bewertet, weil der Mindestlohn nicht zwischen beiden Berufsgruppen differenziert (vgl. IAW 2011c, S. 6). Allerdings ist es möglich, dass aufgrund der gestiegenen relativen Kosten einer Pflegehilfskraft diese durch eine Pflegefachkraft substituiert wird. Dieser Sachverhalt sollte durch einen einfachen Abgleich der Hilfs- und Fachkräftequoten in den Einrichtungen vor und nach der Mindestlohneinführung zu untersuchen sein.

Die Evaluation zur Abfallwirtschaft definiert Betriebe als Treatmentgruppe, die vor Mindestlohneinführung mindestens einen Arbeitnehmer unterhalb des Mindestlohns entlohnten. Als Kontrollgruppe werden die Betriebe gewählt, die keinem Mitarbeiter in Folge des Mindestlohns eine Lohnerrhöhung zahlen mussten, also alle Betriebe, die ihre Mitarbeiter ausnahmslos oberhalb des Mindestlohns entlohnten.

Für die Abfallwirtschaft wurden mögliche Spillovereffekte als wenig relevant angesehen, da das Beschäftigungswachstum in der Kontrollgruppe niedriger ausfiel als in der Treatmentgruppe. Substitutionseffekte sind hier vor allem zwischen regulär Beschäftigten und Teilzeitbeschäftigten sowie Beschäftigten des zweiten Arbeitsmarktes (1-Euro-Jobber) zu erwarten. In dem Gutachten wurde dies deskriptiv untersucht. Es zeigt sich das der Anteil von 1-Euro Jobbern sehr gering ist. Der Anteil an Teilzeitbeschäftigten ist in Unternehmen, die den Mindestlohn umgehen, höher (vgl. ZEW 2011b, Tabelle 33). Aufgrund der Datenlage konnten diese Effekte im Rahmen der Studie nicht weitergehend untersucht werden, sollte aber in zukünftigen Analysen stärker berücksichtigt werden.

2.3.2 Analyseebene

Der DvD-Ansatz wird in beiden Studien ausschließlich auf Unternehmensebene durchgeführt. Im Rahmen der Evaluation des ZEW werden die auf Betriebs-ebene geschätzten Effekte durch Umrechnungsverfahren auf die Branchen-ebene übertragen. Hierbei verwendeten die Autoren die inverse Auswahl-wahrscheinlichkeit der Beobachtungen, also das Verhältnis der Zahl der Un-

ternehmen in der Grundgesamtheit zu den Beobachtungen in einer Zelle und gewichteten diese mit den Stichprobenergebnissen um die Repräsentativität für die Grundgesamtheit zu erreichen.

Zur Prüfung der Robustheit der Ergebnisse verwendeten die Autoren restriktivere Spezifikationen. Unter Beachtung der Tatsache, dass ein Teil der betroffenen Unternehmen trotz Mindestlohneinführung weiterhin unterhalb des Mindestlohns entlohnt, wurde die Treatmentgruppe auf diejenigen Unternehmen eingeschränkt, die nach der Mindestlohneinführung nicht mehr unterhalb von 8 Euro entlohnten. Diese Prüfung wurde mit dem für die Kausalanalyse verwendeten Datensatz durchgeführt und veränderte die Ergebnisse nicht.

Die Evaluation der Pflegebranche untersucht die Effekte der Mindestlohneinführung auf Betriebsebene. Diejenigen Einrichtungen, die zum Zeitpunkt der Mindestlohneinführung tarifgebunden waren, bilden die Kontrollgruppe. Es wird also zwischen Treatment- und Kontrollgruppe weder auf Branchenebene, noch auf Individualebene unterschieden. Die einzige Abgrenzung der Effekte erfolgt im Hinblick auf die Tarifbindung der einzelnen Einrichtungen. Hierbei wäre das Problem der Endogenität der Tarifbindung zu diskutieren gewesen.

2.3.3 Funktionale Form der Zielgröße

Die zu untersuchenden Zielgrößen der Evaluation betreffen die Beschäftigung, den Arbeitnehmerschutz und den Wettbewerb. Während die Veränderung der Beschäftigung als abhängige Variable im Logarithmus angegeben wird, berechnet sich der Arbeitnehmerschutz in der Abfallbranche als Indikatorvariable mit Hilfe eines Probitmodells. Auch für die Variablen, die den Wettbewerb widerspiegeln, wie die Neugründungen und Schließungen von Betrieben, werden Probitmodelle geschätzt (mit Ausnahme der Umsatzentwicklung).

Um die Abgrenzung zwischen Treatment- und Kontrollgruppe zu überprüfen, wurden Robustheitsprüfungen im Rahmen der Evaluation zum Mindestlohn in der Pflegebranche durchgeführt. Hierbei ist wichtig, dass die funktionale Form der Zielgröße dieselbe Skalierung hat wie in den Placebo-Experimenten, da sonst die Annahme paralleler Trends verletzt sein kann. Im DvD-Ansatz wurde eine kontinuierliche, unbeschränkte Ergebnisvariable verwendet, die logarithmiert wurde.

Auf Basis der administrativen Individualdaten wurde der Effekt auf die Ergebnisvariable geschätzt, die jedoch anders konstruiert wurde (vgl. IAW 2011c, S. 169). Die durchgeführten Robustheitsanalysen sind deshalb auch im Hinblick auf die unterschiedliche Skalierung nicht eins-zu-eins vergleichbar.

2.3.4 Datengenauigkeit

In beiden Studien waren amtliche Daten nicht oder nur beschränkt verfügbar. Für die Pflegebranche lagen keinerlei amtliche Daten vor, die den Zeitraum nach Mindestlohneinführung abdeckte, so dass der Auftraggeber eine eigene Befragung schon vor der Vergabe der eigentlichen Evaluation in Auftrag gab. Durch ein Befragungsunternehmen wurde eine Online-Befragung durchgeführt, die einmal den Zeitraum kurz vor der Mindestlohneinführung und den Zeitraum kurz danach abdecken sollte. Der Vorteil dieser Datengewinnung liegt in der recht präzisen Abfrage der Stundenlöhne (in engen Klassen mit einer Klassenbreite von 50 Cent) und der damit verbundenen Vermeidung von Verzerrungen durch die Berechnung aus anderen Datensätzen. Des Weiteren ist eine gute Abgrenzung der vom Mindestlohn betroffenen Einrichtungen und Personen möglich, da die Befragung nicht nur Fragen zur Lohnstruktur sondern auch zu den Einrichtungen und dem Arbeitsumfeld umfasste.

Die gesetzliche Definition für den betrieblichen bzw. den persönlichen Geltungsbereich des Mindestlohns ist allerdings sehr undeutlich gehalten. Nach § 1 der Pflegearbeitsbedingungsverordnung umfasst der Geltungsbereich „betriebliche und selbstständige Betriebsabteilungen, die überwiegend ambulante, teilstationäre oder stationäre Pflegeleistungen für Pflegebedürftige“ nach SGB XI anbieten und schließt Betriebe aus, die „überwiegend ambulante Krankenpflegeleistungen für Pflegebedürftige erbringen“. Ausgeschlossen aus der Kategorie der Pflegebetriebe werden hier „Einrichtungen, in denen die Leistungen zur medizinischen Vorsorge, zur medizinischen Rehabilitation, zur Erziehung kranker oder behinderter Menschen im Vordergrund des Zwecks der Einrichtung stehen, sowie Krankenhäuser“. Dennoch kann vermutet werden, dass die in der Stichprobe aufgenommenen Einrichtungen nach dieser Definition mindestlohnpflichtig sind, da sie anhand der Pflegestatistik und über das Branchenverzeichnis der Deutschen Telekom ermittelt wurden. Es ist darauf hinzuweisen, dass es sich um eine freiwillige Befragung handelte. Trotz der Vorteile dieser Datengrundlage hängt die Aussagekraft der Ergebnisse stark

von der Teilnahmebereitschaft der einzelnen Einrichtungen ab. Systematische Antwortverweigerungen können beispielsweise durch hohe zeitliche Belastung kleinerer Einrichtungen vorliegen oder durch Tarifbindung der Einrichtungen, die somit keinen Nutzen in einer Mindestlohnbefragung sehen.

Amtliche Daten wurden lediglich zur Robustheitsprüfung verwendet. Hier wurden die Stundenlöhne von Vollzeitbeschäftigten aus den IEB mit Hilfe von Arbeitszeiten berechnet, die aus dem Mikrozensus auf Basis der Tagesendgelte ermittelt wurden. Die in der Evaluation berechneten Stundenlöhne sind in Lohngruppen aufgeteilt. Diese unterschiedliche Skalierung der Stundenlöhne erlaubt keinen direkten Vergleich der Koeffizienten in der Robustheitsanalyse.

Die Datengrundlage für die Evaluation der Abfallwirtschaft wies ebenfalls das Problem der mangelnden amtlichen Daten auf. Auch hier wurde eigens für die Evaluation eine Befragung durchgeführt. In diesem Fall setzte sich die Befragung aus 10 ausführlichen, leitfadengestützten Interviews und einer telefonischen Unternehmensbefragung mit 492 Unternehmen der Abfallwirtschaft und 260 Unternehmen der Kontrollbranche zusammen. Ergänzend zu den amtlichen Statistiken verwendeten die Autoren das Mannheimer Unternehmenspanel (MUP), das vom ZEW in Zusammenarbeit mit Creditreform als Panneldatenbank über nahezu alle wirtschaftsaktiven Unternehmen in Deutschland erstellt wird. Vorteilhaft ist das Vorliegen von Branchenklassifikationen nach den verschiedenen Wirtschaftszweigklassifikationen (WZ08 seit 2009, WZ03, WZ93 und WZ79) Außerdem sind seit 1999 jährliche Informationen zu Umsatz, Mitarbeiteranzahl und Bonität enthalten.

Nicht eindeutig ist aber vor allem auch in dieser Branche die Umschreibung der Wirtschaftszweigklassifikation von 2003 (WZ03) auf die Klassifikation im Jahre 2008 (WZ08). Nach der WZ08 fallen zum Teil Betriebe in die Sparte der Abfallentsorgung, die nach der WZ03 zum verarbeitenden Gewerbe oder zur Energieversorgung gehörten. Außerdem wird die Straßenreinigung in der WZ08 sonstigen wirtschaftlichen Dienstleistungen zugeordnet, obwohl sie in der WZ03 noch der Abfallentsorgung angehörte. Durch diese Ungenauigkeiten ist es problematisch, die vom Mindestlohn betroffenen Betriebe eindeutig zu identifizieren. Für die Kontrollbranche existiert dieses Problem nicht.

Eine weitere Datenbasis bildet die 2008 durch das ZEW durchgeführte Befragung der Abfallwirtschaft. Anhand dieser Befragungsdaten konnte der Mess-

fehler der Retrospektivinformationen der Befragung abgeschätzt und berücksichtigt werden. Die Autoren kamen zu dem Schluss, dass der Messfehler die geschätzten Wirkungen der Mindestlohnregelungen nicht systematisch verzerrt. Der Messfehler führt zu höheren Standardfehlern und somit zu einer größeren Ungenauigkeit der geschätzten Koeffizienten (vgl. ZEW 2011b, Kapitel 7.3.4 zur Diskussion der Folgen des Messfehlers auf die Ergebnisse).

2.3.5 Zusammenfassung

Die Evaluationen der Sektoren mit einer spezifischen Wettbewerbssituation verwenden selbst erhobene Befragungsdaten. Die Qualität dieser Befragungsdaten ist hinreichend gut um Wirkungsanalysen der Mindestlohneinführung umzusetzen. Die Studie zum Pflegesektor umfasst eine sehr sorgfältige Analyse der Validität der verwendeten Daten.

Beide Studien fokussieren die Wirkungsanalyse hauptsächlich auf die Betriebs- bzw. Unternehmensebene. In der Pflegebranche wurde nur eine interne Kontrollgruppe verwendet. Substitutionseffekte über eine Verschiebung der Nachfrage nach Pflegedienstleistungen sind hier zu erwarten. Die Autoren der Studie präsentieren eine Robustheitsanalyse, die aufgrund der Datenlage jedoch nicht mit den Hauptergebnissen vergleichbar ist.

Die Zuteilung der Betriebe in die Vergleichsgruppen erfolgt in der Pflegebranche im Hinblick auf deren Tarifbindung. Dabei muss beachtet werden, dass die Zuordnung zur Treatment- und Kontrollgruppe nicht eindeutig sein muss, da eine Tarifbindung auch freiwillig angewendet worden sein kann, indem sie beispielsweise in einem individuellen Arbeitsvertrag vereinbart wurde.

Es besteht weiterhin eine mögliche Endogenität der Tarifbindung selbst. Die Lohnhöhe und die freiwillige Tarifbindung können von der individuellen Leistungsfähigkeit des Betriebes abhängen. Das Branchenbild zeigt, dass die Tarifbindung vor allem vom Anbietertyp anhängig ist. Öffentlich-rechtliche sowie freigemeinnützige und kirchliche Anbieter zahlen weitaus häufiger kollektivrechtlich vereinbarte Löhne als private Träger. Die Vergleichbarkeit mit privaten Anbietern, die in stärkerem Wettbewerb stehen und die keine Tarifbindung aufweisen, ist daher fraglich.

Die Evaluation zur Abfallwirtschaft basiert die Wirkungsanalyse auf einen Vergleich mit Betrieben aus dem Garten- und Landschaftsbau. Die Auswahl dieser

Kontrollgruppe ist nicht ausreichend für die gesamte Branche der Abfallwirtschaft. Die wichtige Annahme paralleler Trends ist in vielen Punkten verletzt und es scheint fraglich, ob sich die Kontrollgruppe auch für den Teilbereich der Abfallbeseitigung eignet. Die gewählte Kontrollgruppe innerhalb der Abfallwirtschaft erfüllt die Annahme paralleler Trends. Ob die Exogenitätsannahme erfüllt ist, müsste im Rahmen der Studie tiefergehend diskutiert und untersucht werden.

Durch die Verwendung von Befragungsdaten konnten beide Studien wichtige Kennzahlen wie die Stundenlöhne in benötigter Form abfragen, so dass keine Verzerrungen durch die Berechnung aus anderen Datensätzen entstehen konnten. Im Hinblick auf die externe Validität der Ergebnisse muss erwähnt werden, dass die verwendeten Daten nicht repräsentativ für die Grundgesamtheit sind. Ein Abgleich mit amtlichen Daten, das Durchführen von vergleichbaren Robustheitsanalysen und die Ermittlung von Effekten auf Individualebene ist ratsam, wenn in zukünftigen Evaluationen eine interne und externe Validität angestrebt werden soll.

2.4 Fazit

Die dargestellte Systematisierung und Bewertung der vorliegenden Evaluationsstudien sollte die grundlegenden Vorgehensweisen, methodische Herausforderungen und Ergebnisse herausarbeiten und bewerten. Über die Bewertungen der einzelnen Branchen hinaus lassen sich folgende Punkte an dieser Stelle als Fazit festhalten. Bei aller Kritik im Folgenden ist darauf hinzuweisen, dass die Evaluationsstudien für die Erstellung der Gutachten insgesamt einen sehr engen Zeitrahmen hatten, der es nicht immer erlaubte, alles wissenschaftlich Wünschenswerte umzusetzen (siehe hierzu auch die methodischen Vorschläge in den folgenden Kapiteln).

1. Alle Mindestlohnstudien haben ein Kontrollgruppendesign verwendet und in dem sehr kurzen Zeitrahmen, der für die Erstellung der Studien zur Verfügung stand, sehr sorgfältige Datenarbeit geleistet und dabei eine Vielzahl von wichtigen Erkenntnissen geliefert. In den meisten Studien wurde ein Differenz-von-Differenzen Ansatz gewählt und sehr sorgfältig die zugrundeliegenden Annahmen überprüft (gemeinsamer Trend, Placebo-Test, Abwesenheit von Spillo-

vereffekten bzw. SUTVA-Annahme). In dem Gutachten für die Gebäudereinigung (IAQ) sind an dieser Stelle jedoch methodische Schwächen zu konstatieren.

2. Aus der Systematisierung und Bewertung der Studien geht hervor, dass die Qualität der Studien stark mit der Verfügbarkeit von administrativen Arbeitgeber-Arbeitnehmerdaten zusammenhängt. Hier ist für die drei Studien zu den erst in der jüngeren Vergangenheit eingeführten Mindestlöhnen (Wäschereigewerbe (IAQ), Abfallwirtschaft (ZEW), Pflege (IAW)) zu konstatieren, dass solche Daten mit hinreichender zeitlicher Länge nach der Einführung des Mindestlohns zum Zeitpunkt der Erstellung der Studien nicht zur Verfügung standen. Für diese drei Branchen wäre eine erneute ökonomische Evaluation unter Berücksichtigung der Erkenntnisse aus der ersten Evaluationsrunde und unter Berücksichtigung einiger der in diesem Gutachten (in diesem Kapitel und in Kapitel 3) gemachten methodischen Hinweise und Vorschläge sinnvoll. Die schwierige Datenlage zum Zeitpunkt der Erstellung der Studien lassen substantielle Verbesserungen in der Belastbarkeit der Ergebnisse erwarten.

Dies gilt a fortiori, weil jede dieser drei Branchen Besonderheiten aufweist (siehe oben), die eine erneute Evaluation sinnvoll erscheinen lassen. Dabei liefern die Gutachten zu den drei genannten Branchen durchaus wichtige Einsichten und es wurde durch zusätzliche umfangreiche (allerdings auch kostenträchtige) Befragungen versucht, eine Datenbasis für eine ökonomische Evaluation zu schaffen. Das dies trotz großer Anstrengungen nicht in vergleichbarer Qualität wie für die fünf anderen Gutachten gelungen ist, ist den Autoren und Autorinnen der Gutachten nicht anzulasten. Gleichwohl hätten insbesondere Gutachten zur Abfallwirtschaft (ZEW) und Wäscherei (IAQ) – wie oben ausgeführt – an einigen Stellen trotz der beschränkten Datenbasis prinzipiell weitergehende Analysen im Hinblick auf die Durchführung von Placebo-Tests und die Wahl alternativer Kontrollbranchen durchführen können. Hierbei ist allerdings nochmals auf den kurzen Zeitrahmen für die Erstellung der Analysen nach Erhebung der Daten hinzuweisen, der dies vermutlich nicht in umfassender Weise zugelassen hätte.

3. Trotz der Verfügbarkeit von aussagekräftigen administrativen Arbeitgeber-Arbeitsnehmerdaten besteht in vielen Bereichen ein starkes Abgrenzungsproblem der von Mindestlohnregelung betroffenen Beschäftigten. Dies hängt vor allem damit zusammen, dass Stundenlöhne in der Beschäftigtenstatistik in der Vergangenheit nicht vorlagen und dass die Branchenzugehörigkeit – u.a. angesichts von Änderungen der Wirtschaftszweigklassifikationen – nicht immer eindeutig bestimmbar ist. Schließlich stellt sich sogar die Frage, ob Betriebe die Branchenzugehörigkeit manipulieren können, um allfälligen Mindestlohnregelungen zu entgehen.
4. In den vier Gutachten für das Bauhaupt- und die Baunebengewerbe wurden in vorbildlicher Form die Voraussetzungen für die DvD-Analyse und die damit zusammenhängende Wahl der Kontrollgruppen untersucht. Allerdings gingen die Gutachten nicht soweit, dass sie verschiedene Kontrollgruppen zusammenfassten (synthetische Kontrollgruppen, siehe Abschnitt 3.1.4), um die Vergleichbarkeit zu verbessern und dass sie bei Abweichungen von den Voraussetzungen für die DvD-Analyse (Beispiel: Spillovereffekte und SUTVA-Annahme) weitergehende Analysen durchführten, um diese Abweichungen zu interpretieren. Einige Studien sahen in der Robustheit von Ergebnissen gegenüber der Wahl der Kontrollgruppe ein anstrebenswertes Ergebnis. Umgekehrt könnten sich jedoch aus der Variation der Ergebnisse je nach Wahl der Kontrollgruppen neue Erkenntnisse ergeben, sprich diese Unterschiede könnten per se Anhaltspunkte für die ökonomische Wirkungsweise der Mindestlöhne geben (bspw. weil die verschiedenen Kontrollgruppen von unterschiedlichen Spillovereffekten betroffen sein könnten, die explizit zu untersuchen wären). Dies könnte jedoch auch Anlass für die Verwendung synthetischer Kontrollgruppen sein, die in optimaler Weise den Vergleichsmaßstab für die vom Mindestlohn Betroffenen auf Basis verschiedener Kontrollgruppen schätzt. In den hier genannten Punkten besteht Verbesserungsbedarf in der nächsten Evaluationsrunde (siehe die Ausführungen in den folgenden Kapiteln).

5. Aufgrund der Spezifika der einzelnen Branchen und der unterschiedlichen Höhe der Mindestlöhne ist die externe Validität der Ergebnisse sehr begrenzt. Die Evaluationsstudien erlauben nur sehr beschränkte Aussagen über die Einführung von Mindestlöhnen in anderen Branchen.
6. Ein in allen Studien zu wenig untersuchtes Problem ist die mögliche Politikendogenität der Einführung oder Veränderung von Mindestlohnregelungen. Dieses Problem kann die Kontrollgruppendesigns in Frage stellen (siehe oben) und wäre genauer zu untersuchen. Ob ein Problem der Politikendogenität vorliegt, kann zunächst durch spezifische qualitative Befragungen der Akteure in einer Branche bestimmt werden. Das Problem kann sich auch darin widerspiegeln, dass für bestimmte gewählte Kontrollgruppen Placebo-Tests scheitern und dies auf eine andere erwartete zukünftige Entwicklung für die Betroffenen ohne Änderung der Mindestlohnregelungen in Vergleich zu den Kontrollgruppen hinweist. Es wäre dann zu prüfen, ob die Extrapolation der geschätzten Trends mit den Erwartungen der Akteure in den Branchen übereinstimmen (hier könnten auch Informationen aus regelmäßigen Unternehmensbefragungen – Beispiele sind der ifo Konjunkturtest oder die Konjunkturumfrage des DIHK – über die Erwartungen verwendet werden), um ein Gesamtbild zu erhalten. Die geschätzten Effekte der Änderungen der Mindestlohnregelungen im Vergleich zu einer/mehreren Kontrollgruppen könnten im Lichte dieser Trendextrapolationen und plausibler Annahmen über die Reaktionen der Unternehmen (bspw. mit stärker strukturellen Analysen, siehe Kapitel 4) korrigiert werden.
7. Die Studie zum Bauhauptgewerbe (IAB, RWI, ISG), die in vielerlei Hinsicht vorbildlich ist, verwendet den von Dolton et al. (2010, 2012a) vorgeschlagenen inkrementellen DvD-Ansatz, der aufgrund des oben Gesagten und der Ausführungen in Kapitel 3.1.3 sehr kritisch im Hinblick auf die Messung der Wirkungen über mehrere Jahre bei sich ändernder Eingriffsintensität einzuschätzen ist. Sinnvoll ist der Ansatz jedoch im Hinblick darauf, dass für die Einschätzung der Wirkung von Änderungen der Mindestlohnregelungen zwischen

zwei Zeitpunkten die Wirkungsunterschiede je nach Eingriffsintensität erfasst werden können.

8. An mehreren Stellen in den Studien deutet sich an, dass Betriebe in einer Weise reagieren, die nicht der Intention der Mindestlohnregulierung entspricht (Beispiele: unbezahlte Stunden, Arbeitsverdichtung) oder die zumindest den als positiv eingeschätzten Wirkungen des Mindestlohns gegenüber gestellt werden müssten (Beispiele: Überwälzung der erhöhten Lohnkosten auf die Kunden, geringere Fluktuation der Beschäftigten, Rückgang der Lohndifferenzierung, Zunahme von „Recalls“). Hier sind auch Themen wie die mögliche Zunahme atypischer Beschäftigung oder Schwarzarbeit zu nennen. Angesichts der Vielschichtigkeit der möglichen Reaktionen wäre ein umfassenderes Untersuchungsdesign sinnvoll, das anstelle der getrennten Schätzung der Effekte auf einzelne Ergebnisvariablen den Blick auf das Zusammenspiel der Wirkungen (oben hatten wir schon auf das Zusammenspiel der Lohn- und Beschäftigungswirkungen hingewiesen, die für die Einschätzung von Spillover- und Substitutionseffekten gemeinsam untersucht werden müssen) oder die gesamte Anpassungsstrategie legt. Soweit diese Effekte gemeinsam beobachtet werden können, ist eine solche umfassendere Wirkungsanalyse mit ökonometrischen Methoden möglich.
9. Es wäre sinnvoll gewesen, in noch stärkerem Umfang ein einheitliches Forschungsdesign für alle Branchen (zumindest für die fünf Branchen, in denen Mindestlohnregelungen schon seit längerer Zeit vorliegen) anzuwenden. Ansatzweise ist dies gegeben, da das IAW den Evaluationsauftrag für drei Branchen hatte. Ein einheitliches Forschungsdesign könnte die externe Validität der erzielten Ergebnisse erhöhen, da dann in einem Evaluationsgutachten die Unterschiedlichkeit der erzielten Ergebnisse „erklärt“ werden müsste.

3 Alternative methodische Ansätze zur Ex-Post-Evaluation (Nicht-Strukturelle Modelle)

In den folgenden Abschnitten sollen zunächst alternative methodische Ansätze zur Ex-Post-Evaluation, zu denen auch der DvD-Ansatz zählt, diskutiert werden. Im Rahmen einer Ex-Post-Evaluation werden die Folgen einer tatsächlich beobachteten Einführung oder Veränderung von Mindestlöhnen untersucht. Der Fokus liegt somit auf einer bestimmten Politikvariation, deren Wirkungen auf die zentralen Ergebnisvariablen der Politik, etwa die Beschäftigung in einer Branche quantifiziert werden sollen. Anders als bei strukturellen Ansätzen wird auf eine Modellierung der generellen Wirkungszusammenhänge verzichtet. In Abschnitt 3.1 erfolgt zunächst eine Diskussion verschiedener Varianten des DvD-Ansatzes, während Abschnitt 3.2 alternative Methoden zur Ex-Post-Evaluation in den Blick nimmt.

3.1 Varianten des DvD-Ansatzes

3.1.1 DvD kombiniert mit Matching

Wie bereits in Kapitel 2 dargelegt wurde, besteht eine zentrale Annahme bei herkömmlichen DvD-Ansätzen darin, dass sich die Erfolgsgrößen der Treatment- und Kontrollgruppe gleich entwickelt hätten, hätte es den Mindestlohn nicht gegeben (Stable Unit Treatment Value Assumption, SUTVA). Diese Annahme ist in der Praxis oft nicht erfüllt, da perfekt vergleichbare Kontrollgruppen meist nicht vorhanden sind (z.B. Beschäftigte der gleichen Branche ohne Mindestlohn). In der Folge weisen Treatment- und Kontrollgruppe bereits vor der Politikintervention systematische Unterschiede auf (Selektionsverzerrung). Eine Möglichkeit diese zu reduzieren besteht darin, die Vergleichbarkeit der Kontrollgruppe mit Hilfe eines kombinierten DvD-Ansatzes mit Matching zu erhöhen (Heckman et al. 1997, Dehejia und Wahba 1999). Die Idee des nicht-parametrischen Verfahrens besteht darin, eine neue vergleichbare Kontrollgruppe zu generieren, indem zu jedem Individuum in der Treatmentgruppe ein statistischer Zwilling aus der Kontrollgruppe gesucht wird, der möglichst ähnlich in seinen sozio-ökonomischen Merkmalen ist, nicht aber von der Maßnahme betroffen ist. Hierbei sollten nur Merkmale berücksichtigt werden, die die Ergebnisvariable potenziell erklären können und nicht die Ergebnisvariable

selbst. Wurde jedem Beschäftigten eine vergleichbare Person aus den Daten zugeordnet (gematcht), kann der DvD Schätzer mit Hilfe der Treatment- und der neu gezogenen Kontrollgruppe durchgeführt werden.

Ein exaktes Matching ist ein sehr datenintensives Verfahren, da alle Merkmale beobachtet werden müssen, die gleichzeitig die Betroffenheit vom Mindestlohn (Teilnahmewahrscheinlichkeit) und die Zielgröße beeinflussen. Je mehr Merkmale bzw. Variablen man aufnimmt, desto schwieriger ist es, eine Beobachtung im Datensatz mit genau den gleichen Eigenschaften wie etwa Berufserfahrung, Alter, Bildung, Arbeitsmarkthistorie oder Familienstand zu finden. Wie Rosenbaum und Rubin (1983) allerdings zeigen, ist es ausreichend, Teilnehmer und Nichtteilnehmer auf Basis ihrer geschätzten Teilnahmewahrscheinlichkeit zuzuordnen, was die Dimensionalität stark reduziert. Bei dem sogenannten Propensity Score Matching wird zunächst die Wahrscheinlichkeit geschätzt, von der Maßnahme betroffen zu sein, konditional auf sozio-ökonomische Merkmale. Die vorhergesagte Betroffenheitswahrscheinlichkeit bildet dann den Propensity Score, der den Einfluss der für die Selektion relevanten Charakteristika zusammenfasst. Anhand des Propensity Scores wird dann einer Person mit Treatment eine Vergleichsperson ohne Treatment mit ähnlichem Propensity-Score zugewiesen. Der Unterschied in den Erfolgsgrößen (nach Treatment) ist der kausale Effekt des Treatments.

Formal wird die Differenz der Veränderung der Ergebnisgröße Y zwischen Teilnehmer i und zugeordneten Kontrollpersonen j vor (t_0) und nach (t_1) der Maßnahme geschätzt. Der ATT (Average Treatment Effect on the Treated) ergibt sich somit als:

$$ATT^{DvD} = \sum_{i \in Treatment} \left((Y_{it_1} - Y_{it_0}) - \sum_{j \in Kontroll} g(\hat{p}_i, \hat{p}_j) (Y_{jt_1} - Y_{jt_0}) \right),$$

wobei der Term $g(\hat{p}_i, \hat{p}_j)$ den Gewichtungsfaktor der Nichtteilnehmer angibt (siehe Blundell und Costa Dias 2009 für eine ausführliche Darstellung). $g(\hat{p}_i, \hat{p}_j)$ bemisst sich nach der Ähnlichkeit zur Treatmentgruppe auf Basis der berechneten Propensity Scores \hat{p}_i und \hat{p}_j . Für die Zuordnung können verschiedene Algorithmen benutzt werden, wie etwa das Nearest-Neighbour-Matching oder Kernel-based Matching Verfahren (für eine detaillierte praktische Anleitung siehe Caliendo und Kopeinig (2008)).

Eine zentrale Annahme beim Matching ist die Conditional Independence Assumption (CIA) bzw. die Annahme der Unconfoundedness, die verlangt, dass Unterschiede in der Erfolgsgröße einer Person mit Treatment und dem Ergebnis einer anderen, bezüglich X äquivalenten Person ohne Treatment (der statistische Zwilling der Person), ausschließlich auf das Treatment zurückgehen und nicht auf die Selektion ins Treatment. Unbeobachtete Unterschiede zwischen den beiden Personen können allerdings immer noch bewirken, dass sich die Erfolgsgrößen unterscheiden und damit die Annahme verletzt ist. Dieses Problem kann dadurch vermindert werden, dass beispielsweise auch unbeobachtete Charakteristika wie Motivation durch Informationen zur Arbeitsmarkthistorie oder ähnliches approximiert werden.

Eine weitere Annahme des Matching verlangt, dass genügend Individuen mit ähnlichen Charakteristika einander zugeordnet werden können (Common Support). Mit anderen Worten die Verteilung des Propensity Scores für die Treatment und Kontrollgruppe muss sich hinreichend überlappen. Unter der üblichen Annahme, dass sich die Treatment und Kontrollgruppe ähnlich entwickelt hätten, wäre der Mindestlohn nicht eingeführt worden (Common Trends) kann der konditionale und unkonditionale durchschnittliche Treatment Effekt der betroffenen Beschäftigten (ATT) identifiziert werden.

Im Vergleich zu einem konditionalen DvD-Schätzer, d.h. einer parametrischen DvD-Schätzung unter Verwendung beobachteter Charakteristika, sind für das nicht-parametrische Matching-Verfahren jedoch größere Datenmengen erforderlich. Zudem kann es schnell zu weiteren Verzerrungen kommen, wenn die Matching-Variablen nicht sorgfältig ausgewählt werden (Heckman und Navarro 2004). Auch die Interpretation der Parameter ist unklar, wenn nicht alle von der Maßnahme betroffenen Observationen gematcht werden können (Blundell und Dias 2009). Weiterhin können Antizipationseffekte die Methodik in Frage stellen. Der Umgang mit Antizipationseffekten und mögliche Lösungen werden in Bergemann et al. (2009) detailliert diskutiert.

Eine Kombination aus Matching und DvD wurde bereits vielfach für Evaluationen arbeitsmarktpolitischer Maßnahmen angewendet (für einen Überblick siehe Kluge (2010)). Mit Ausnahme der Studie von Nguyen (2013) fand die Methodik allerdings kaum Anwendung in der Mindestlohnforschung.

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne: In den bestehenden Studien der ersten Evaluationsrunde deutscher Branchenmindestlöhne wurden vielfach konditionale DvD-Ansätze verwendet, d.h. die Differenz-von-Differenzen wurde konditional auf beobachtete Faktoren bestimmt, um eventuelle, ergebnisrelevante Unterschiede zwischen der Treatment- und Kontrollgruppe herauszurechnen. Einer Selektion in die Treatmentgruppe auf Basis beobachteter Merkmale wird hier somit Rechnung getragen, wobei der Zusammenhang parametrisch geschätzt wird und somit Annahmen bzgl. des funktionalen Zusammenhangs notwendig sind. Auch das Matching-Verfahren kombiniert mit DvD ist ein Versuch, eine Selektionsverzerrung aufgrund beobachteter Charakteristika zu vermeiden. Im Gegensatz zum parametrischen DvD-Schätzer, handelt es hierbei jedoch um ein flexibleres, nicht-parametrisches Verfahren. DvD mit Matching bzw. ein konditionaler DvD-Schätzer berücksichtigen jedoch dieselben Ursachen einer Selektionsverzerrung: beide eliminieren Unterschiede in beobachteten Charakteristika und differenzieren zeitinvariante, unbeobachtete Faktoren heraus. Je mehr Informationen über die Beschäftigten somit vorhanden sind, umso eher kann eine Selektionsverzerrung minimiert werden. Die üblichen Annahmen des DvD-Ansatzes – Common Trends und SUTVA – müssen jedoch in beiden Fällen weiterhin gelten. Je vergleichbarer Treatment- und Kontrollgruppe sind, umso eher sind diese Annahmen plausibel, müssen aber auch dann nicht zwingend gelten.

Insbesondere gibt es für die möglichen Kontrollgruppen im Fall der Branchenmindestlöhne Gründe, die Gültigkeit der Annahmen weiterhin zu hinterfragen. Bei einem Vergleich verschiedener Branchen kann zwar durch ein Matching-Verfahren die Vergleichbarkeit von Beschäftigten bzw. Betrieben im Hinblick auf beobachtete Charakteristika gewährleistet werden, so dass die Branchenzugehörigkeit idealerweise wie ein Zufallsexperiment betrachtet werden kann. Das Problem der Politikendogenität wird dadurch hingegen nicht gelöst. Insbesondere gewährleistet die statistische Vergleichbarkeit von Treatment- und Kontrollgruppe nicht automatisch, dass die Annahme paralleler Trends gültig ist. Vielmehr wäre es selbst für ein und dasselbe Individuum in der vom Mindestlohn betroffenen und in einer nicht vom Mindestlohn betroffenen Branche denkbar, dass sich der Erwerbsverlauf unterschiedlich entwickeln würde, weil sich die Branchen unterschiedlich entwickeln. Dieses Problem könnte

durch die Anwendung einer synthetischen Kontrollgruppe abgemildert werden (vgl. hierzu Abschnitt 3.1.4). Zwar löst auch der synthetische Kontrollgruppenansatz nicht das Problem der Politikendogenität, aber zumindest stellt dieser Ansatz eine Vergleichbarkeit für die Zeit vor der Mindestlohneinführung auf Branchenebene her. Die Common Trends Annahme ist somit eher gültig als im Falle eines Verfahrens, welches lediglich auf die Vergleichbarkeit von Beschäftigten oder Betrieben abzielt.

Im Fall eines Matchings innerhalb der Branche, d.h. die Identifikation statistischer Zwillinge aus der Menge aller Beschäftigten mit einem Durchschnittslohn oberhalb und unterhalb des Mindestlohns, lässt sich bereits die Grundidee des Matchings hinterfragen. Denn ein Matching-Verfahren lebt von der Idee, dass es konditional auf alle relevanten beobachteten Faktoren einen Zufallsfaktor geben muss, der dann den Unterschied zwischen Treatment und Nicht-Treatment ausmacht. Es müsste somit gelingen, Individuen zu finden, die konditional auf alle Faktoren, welche vor allem die Produktivität des Beschäftigten (bzw. Betriebs) beschreiben, unterschiedliche Lohnniveaus verdienen. Unterschiedliche Löhne konditional auf die Produktivität sind jedoch nur für unterschiedliche Teilmärkte wie z.B. Ost- und Westdeutschland denkbar. Ob die Zugehörigkeit zu einem Teilmarkt dann tatsächlich als Zufallsexperiment angesehen werden kann, ist fraglich.

Unabhängig davon, ob das Matching branchenübergreifend oder innerhalb der betroffenen Branche stattfindet, können Spillovereffekte und andere Abhängigkeiten zwischen Treatment- und Kontrollgruppe zudem die SUTVA verletzen und zu Verzerrungen der Schätzergebnisse führen.

Insgesamt ist somit zu konstatieren, dass ein DvD mit Matching somit keine substanzielle Verbesserung im Hinblick auf die zentralen Identifikationsprobleme gegenüber einem konditionalen DvD-Schätzer aufweist. Ein gewisser Vorteil des DvD mit Matching besteht sicherlich darin, dass die Common Support Annahme den Blick auf die Vergleichbarkeit der Treatment- und Kontrollgruppe lenkt. Nur wenn es gelingt, statistische Zwillinge zu identifizieren, kann das Matching-Verfahren angewendet werden. Das parametrische Verfahren kann allein aufgrund der funktionalen Form einen Treatment-Effekt schätzen; eine hinreichende Vergleichbarkeit von Treatment- und Kontrollgruppe muss jedoch nicht gewährleistet sein. Auf der anderen Seite ist die nicht-

parametrische Spezifizierung des Matching-Ansatzes datenintensiver und somit in ihrer Anwendbarkeit eventuell eingeschränkter.

Im Falle der Branchenmindestlöhne ist somit festzuhalten, dass ein Matching kombiniert mit DvD grundsätzlich möglich ist, wenn ausreichend Daten für das datenintensivere Matching-Verfahren zur Verfügung stehen. Dies ist vor allem für die Branchen des Typs 1 aus dem Bauhaupt- und Baunebengewerbe der Fall, für die umfangreiche administrative Daten vorhanden sind. Im Vergleich zu einem konditionalen DvD-Schätzer reduziert ein DvD-Schätzer mit Matching die grundlegende Identifikationsproblematik hingegen nicht. Das ist vielleicht auch der Grund, warum die Methodik kaum Anwendung in der Mindestlohnforschung fand.

3.1.2 Changes-in-Changes (CiC)

Der Changes-in-Changes-Ansatz, der von Athey und Imbens (2006) entwickelt wurde, ist eine Verallgemeinerung des DvD-Ansatzes (siehe Lechner 2011 für einen Überblick). Letzterer kann als Spezialfall des Changes-in-Changes-Ansatzes interpretiert werden, der auf zusätzlichen Annahmen beruht und daher restriktiver ist.

Ein wesentlicher Vorteil des Changes-in-Changes-Ansatzes ist es, dass die zugrundeliegenden Annahmen des Ansatzes unabhängig von der Skalierung der Ergebnisvariablen sind. Dies trifft für die im DvD-Ansatz relevante Common-Trends-Annahme nicht zu. Diese unterstellt, dass die Entwicklung der Ergebnisvariablen in der Treatment- und Kontrollgruppe identisch gewesen wäre, wenn das Treatment nicht stattgefunden hätte (siehe hierzu Abschnitt 2.1.3).

Die Plausibilität dieser Annahme hängt von der Skalierung der Ergebnisvariablen ab. Wenn die Annahme für eine Ergebnisvariable (z.B. den Lohn) als Niveaugröße Gültigkeit besitzt, muss sie nicht für den Logarithmus dieser Variable gelten. Die Ergebnisse einer Kausalanalyse mittels eines DvD-Ansatzes können dann nur für eine in Niveaugrößen gemessene Ergebnisvariable belastbar interpretiert werden. Der Changes-in-Changes-Ansatz lässt sich hingegen auch bei Transformationen der Niveaugröße anwenden.

Um den Changes-in-Changes-Ansatz im Folgenden formal zu beschreiben und auf Unterschiede zum DvD-Ansatz hinzuweisen, sei angenommen, dass zwei Zeitperioden ($T = 0$ und $T = 1$) und zwei Gruppen ($G = 0$ und $G = 1$) betrachtet

werden. Mit $G = 1$ sei die Treatmentgruppe und mit $G = 0$ die Kontrollgruppe bezeichnet. Die Periode $T = 0$ verweist auf die Zeit vor dem (potenziellen) Treatment und die Periode $T = 1$ auf die Zeit danach. Die interessierende Ergebnisvariable sei mit Y bezeichnet. Unter dieser Notation lässt sich im DvD-Ansatz die Ergebnisvariable eines beliebigen Individuums i für den Fall ohne Treatment, $Y_i(0)$, wie folgt darstellen:¹⁰

$$Y_i(0) = \alpha + \beta \cdot T_i + \gamma \cdot G_i + \varepsilon_i,$$

wobei α , β und γ unbekannte Parameter sind. Der Parameter β identifiziert den Einfluss der Zeit auf die Ergebnisvariable, der unabhängig von der Gruppenzugehörigkeit ist. γ misst hingegen den zeitinvarianten Einfluss der Gruppenzugehörigkeit. α ist eine Konstante, die unabhängig von der Zeit und Gruppe wirkt, und ε_i misst unbeobachtete Charakteristika von Individuum i .

Der DvD-Ansatz nimmt an, dass die Ergebnisvariable additiv linear von einem Zeiteffekt, einem Gruppeneffekt und unbeobachteten Charakteristika abhängt. Dieser additiv lineare Zusammenhang wurde häufig als zu restriktiv für ökonomische Anwendungen kritisiert (Athey und Imbens 2006). Der Changes-in-Changes-Ansatz hebt daher den additiv linearen Zusammenhang auf und modelliert die Ergebnisvariable für den Fall ohne Treatment als

$$Y_i(0) = h_0(U_i, T_i).$$

$h_0(u, t)$ ist eine beliebige monotone Funktion, deren Werte mit steigendem u zunehmen.¹¹ U_i ist eine Zufallsvariable, die die unbeobachteten Charakteristika von Individuum i über den Rang in der statistischen Verteilung von $Y_i(0)$ abbildet. Der funktionale Zusammenhang des Changes-in-Changes-Ansatzes drückt aus, dass in einer Periode alle Individuen mit der Ausprägung $U_i = u$ denselben

¹⁰ Die Darstellung orientiert sich an Imbens und Wooldridge (2009) sowie Athey und Imbens (2006), Lechner (2012) und Melly und Santangelo (2013).

¹¹ Diese Bedingung schließt diskrete Ergebnisvariablen aus, wenn eine stetige Verteilung der unbeobachteten Charakteristika (wie etwa in üblichen diskreten Auswahlmodellen) unterstellt wird. Athey und Imbens (2006, Abschnitt 4) diskutieren, wie der Ansatz auf diskrete Ergebnisvariablen erweitert werden kann. Allerdings ist in diesem Fall nur eine Intervallschätzung (Bounds-Analyse) möglich, was die Anwendung der Methode sicherlich einschränkt.

Wert der Ergebnisvariable aufweisen, unabhängig davon, ob sie in der Treatment- oder Kontrollgruppe sind. Die Verteilung der zufälligen Ränge U_i darf zwischen beiden Gruppen beliebig variieren, aber die Verteilungen der beiden Gruppen müssen sich überlappen (für die Schätzung des Treatmenteffektes für die Treatmentgruppe reicht es aus, wenn der Träger der Verteilung der Ränge für Gruppe $G=1$ innerhalb des Trägers der Verteilung der Ränge für Gruppe $G=0$ liegt, siehe Athey and Imbens (2006), Assumption 3.4). Die Verteilung der Ränge in der Gesamtverteilung darf sich für eine Gruppe nicht über die Zeit ändern (Verteilungsinvarianzannahme). Dies ist die nicht-parametrische Analogie zur Common-Trends-Annahme und sie bietet die Grundlage für Placebo-Tests zur Überprüfung dieser Annahme vor Einführung des Mindestlohns. Für ein einzelnes Individuum i jedoch sind unterschiedliche Ausprägungen von U_i in Periode $T = 0$ und $T = 1$ zugelassen.

Für den Fall ohne Treatment nimmt der Changes-in-Changes-Ansatz an, dass sich alle Unterschiede in der Ergebnisvariable zwischen der Teilnehmer- und Kontrollgruppe aus den Unterschieden in der Verteilung von U ergeben. Der Ansatz nimmt ferner an, dass sich ohne das Treatment Veränderungen in der Ergebnisvariable über die Zeit in beiden Gruppen ausschließlich aus der Tatsache ergeben, dass sich $h_0(u, 0)$ und $h_0(u, 1)$ voneinander unterscheiden, sich der Zusammenhang zwischen den unbeobachteten Charakteristika und der Ergebnisvariablen über die Zeit also verändert.

Die grundlegende Idee des CiC-Ansatzes besteht darin, vier verschiedene Verteilungen der Ergebnisvariable zu betrachten: die Verteilung der Ergebnisvariable für die Treatmentgruppe vor und nach Einführung des Treatments und die Verteilung der Ergebnisvariable für die Kontrollgruppe vor und nach Einführung des Treatments. Die Differenz der Verteilungen der Ergebnisvariablen zwischen Treatment- und Kontrollgruppe vor Einführung des Treatments wird herangezogen, um zu approximieren, wie sich die Ergebnisgrößen in Treatment- und Kontrollgruppe nach Einführung des Treatments unterscheiden würden, hätte es das Treatment hypothetisch nicht gegeben. Der Unterschied zwischen der tatsächlichen Differenz zwischen Treatment- und Kontrollgruppe nach Einführung des Treatments und der approximierten Differenz kann als Effekt des Treatments interpretiert werden. Da die gesamte Verteilung der Ergebnisvariablen betrachtet wird, lässt sich der Effekt des Treatments für jeden beliebigen Punkt oder jedes beliebige Quantil der Verteilung ermitteln.

Liegt das Interesse auf dem durchschnittlichen Effekt des Treatments, θ , für die Treatmentgruppe $G=1$, ergibt sich dieser formal aus

$$\begin{aligned}\Theta &= E[Y_i(1) - Y_i(0) | G_i = 1, T_i = 1] \\ &= E[Y_i(1) | G_i = 1, T_i = 1] - E[Y_i(0) | G_i = 1, T_i = 1],\end{aligned}$$

wobei $Y_i(1)$ die Ergebnisvariable für den Fall des Treatments darstellt und $E[\dots]$ der Erwartungswert-Operator ist.

Der erste Operand der Gleichung, $E[Y_i(1) | G_i = 1, T_i = 1]$, kann direkt aus den Daten ermittelt werden als der Durchschnitt der Ergebnisvariablen in der Treatmentgruppe für die Periode $T = 1$. Der zweite Operand der Gleichung, $E[Y_i(0) | G_i = 1, T_i = 1]$, ist nicht direkt bestimmbar und muss geschätzt werden. Athey und Imbens (2006) zeigen, dass folgender Zusammenhang gilt:

$$E[Y_i(0) | G_i = 1, T_i = 1] = E[F_{01}^{-1}(F_{00}(Y_{i10}))],$$

wobei $F_{gt}(y)$ die Verteilungsfunktion von Y_i gegeben $G_i = g$ und $T_i = t$ ist. $F_{00}(Y_{i10})$ ist der Rang, den Person i in der Treatmentgruppe ($G=1$) in der Verteilung der Ergebnisvariablen der Kontrollgruppe ($G=0$) in Periode 0 hat. Die kontrafaktische Ergebnisvariable für diese Person in Periode 1 wird mit $F_{01}^{-1}(F_{00}(Y_{i10}))$ geschätzt, d.h. es wird unterstellt, dass die Person auch in Periode 1 den gleichen Rang in der Verteilung der Ergebnisvariablen der Kontrollgruppe ($G=0$) hätte. Hier wird besonders deutlich, dass auch der Changes-in-Changes-Ansatz die Gültigkeit der SUTVA-Annahme erfordert und dass sich die Verteilungen für Treatment- und Kontrollgruppe in Periode $T=0$ überlappen müssen.

Die für die Schätzung des Treatmenteffektes notwendigen Verteilungsfunktionen können aus den Daten geschätzt werden, so dass der durchschnittliche Effekt des Treatments auf die Treatmentgruppe im Changes-in-Changes-Ansatz als

$$\theta = E[Y_i(1) | G_i = 1, T_i = 1] - E[F_{01}^{-1}(F_{00}(Y_{i10}))]$$

bestimmbar ist.

Das oben dargestellte Verfahren basiert auf dem Vergleich unkonditionaler Verteilungsfunktionen der Ergebnisvariablen. Das Vorliegen von Heterogenität im Hinblick auf beobachtete Charakteristika kann jedoch insbesondere zur Erfüllung der Überlappungen der Verteilungen der unbeobachteten Charakte-

ristika die Berücksichtigung von Kovariaten erfordern. Hier bietet sich entweder eine getrennte Durchführung des Verfahrens innerhalb von durch diskrete Werte der Kovariaten bestimmten Zellen an (Lechner 2011) oder alternativ der jüngst von Melly und Santangelo (2013) modifizierte Changes-in-Changes-Schätzer für stetige Ergebnisvariablen, der die Berücksichtigung von Kovariaten über die Schätzung von Quantilsregressionen ermöglicht.

Die Anwendung des Changes-in-Changes-Ansatzes für diskrete Ergebnisvariablen (wie die individuelle Beschäftigung) ist möglich, aber methodisch aufwändig. Der Ansatz lässt nur eine Intervallschätzung zu (Bounds-Analyse), die beim Vorliegen von Kovariaten verbessert werden kann (Athey und Imbens 2006, Kapitel 4). Belastbare Anwendungen des Ansatzes liegen für diesen Fall noch nicht vor.

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne: Die Ermittlung von $E[F_{01}^{-1}(F_{00}(Y_{i10}))]$ erfordert eine hohe Anzahl an Beobachtungen und kann mit beträchtlichem rechnerischen Aufwand verbunden sein. Nicht zuletzt deshalb finden sich in der empirischen Wirtschaftsforschung bisher nur sehr wenige Studien, die den Changes-in-Changes-Ansatz verwenden. In der Mindestlohnforschung wurde der Ansatz bisher noch nicht eingesetzt. Da für die Evaluation der Mindestlöhne in Deutschland auf umfassende Datenbasen (bis hin zu Vollerhebungen) zurückgegriffen werden kann, wären die verfügbaren Daten vermutlich kein Hindernis für die Verwendung des Changes-in-Changes-Ansatzes für stetige (oder nahezu stetige) Ergebnisvariablen wie Verdienste oder die Beschäftigung in Betrieben. Da in den bisherigen Studien der DvD-Ansatz dominiert, böte es sich in jedem Fall an, die Robustheit der DvD-Ergebnisse in Bezug auf die Skalierungsabhängigkeit mittels des Changes-in-Changes-Ansatzes zu überprüfen. Ähnlich zum DvD-Verfahren erfordert der Ansatz Daten vor und nach der Mindestlohneinführung und kann sowohl mit wiederholten Querschnittsdaten und Paneldaten durchgeführt werden. Geeignete Placebo-Tests der Verteilungsinvarianzannahme können auf Basis von mehreren Zeitpunkten vor Einführung eines Mindestlohns oder vor einer Veränderung der Mindestlohnregelungen durchgeführt werden. Die praktische Anwendbarkeit des Changes-in-Changes-Ansatzes für diskrete Ergebnisvariablen zur Evaluation der Mindestlöhne in Deutschland kann zum jetzigen Zeitpunkt ex-ante nicht beurteilt werden, da der Ansatz nur eine Intervallschät-

zung ermöglichen würde und hierzu keine belastbaren Anwendungen in der Literatur vorliegen.

3.1.3 Regionale DvD-Ansätze und inkrementelles DvD

In der US-Literatur der 1990er Jahre wurden vielfach Mindestlohnstudien durchgeführt, die zur Identifikation des kausalen Mindestlohneffektes dessen unterschiedliche Höhen auf der Ebene der US-Bundesstaaten ausgenutzt haben. Unter Ausnutzung der regionalen Variation in der Mindestlohnhöhe bzw. ihrer Erhöhungen verwendeten die Studien einen regionalen Panelansatz, der beobachtete und zeitkonstante unbeobachtete regionale Faktoren berücksichtigt und den kausalen Effekt des Mindestlohns unter zwei Annahmen abschätzen kann, die im Wesentlichen den DvD- Annahmen entsprechen: Zum einen wird angenommen, die betrachtete Zielgröße hätte sich in den verschiedenen Regionen konditional auf beobachtete Unterschiede und zeitkonstante unbeobachtete Faktoren ohne den Mindestlohn bzw. die Mindestloohnerhöhung gleich entwickelt. Zum anderen darf es keine Wechselwirkungen zwischen den Regionen in Form z.B. von Spillovereffekten geben. Zudem trägt dieser Ansatz dem Problem der Politikendogenität nicht Rechnung. Dennoch wurde diese Panelmethode vielfach angewendet, vgl. Neumark und Wascher (2006) für einen Überblick.

Im Fall eines flächendeckenden Mindestlohns kann diese Methode nicht unmittelbar angewendet werden, da es keine Variation in der Mindestlohnhöhe gibt. Card (1992) und andere Autoren argumentieren jedoch, dass es auch bei einem überregional einheitlichen Mindestlohniveau eine zu Identifikationszwecken nutzbare regionale Variation in der Betroffenheit vom Mindestlohn gibt. Da sich regionale Märkte strukturell unterscheiden und somit auch Lohnverteilungen regional variieren, unterscheidet sich die Eingriffsintensität eines Mindestlohns zwischen den Regionen. Diese wurde in verschiedenen Studien durch die Betroffenheitsquote (=Anteil der Beschäftigten in einer Region mit einem Lohn unterhalb des Mindestlohns oder am Mindestlohn), den sogenannten Spike (=Anteil der Beschäftigten am Mindestlohn) sowie den Kaitz-Index (=Verhältnis des Mindestlohns zum Durchschnitts- oder Medianlohn) approximiert (Neumark und Wascher 1992, Stewart 2002).

Stewart (2002) schätzt beispielsweise den Zusammenhang zwischen der Beschäftigtenquote auf regionaler Ebene und dem regionalen Anteil an Beschäf-

tigten, die vor Einführung des flächendeckenden Mindestlohns in Großbritannien unterhalb des Mindestlohnlevels verdienen. In der Argumentation dieses Ansatzes ist die Eingriffsintensität konditional auf beobachtete Faktoren und unbeobachtete zeitkonstante Unterschiede exogen und kann somit den Mindestlohneffekt identifizieren. Tatsächlich sollte die Politikendogenität in diesem Fall keine Endogenität des Betroffenheitsindikators induzieren. Es stellt sich jedoch die Frage, ob es im Fall eines überregional einheitlichen Mindestlohnlevels überhaupt eine regionale Variation in der Betroffenheit gibt, wenn für die strukturellen Unterschiede, die zu der unterschiedlichen Betroffenheit führen, hinreichend kontrolliert wird. Anders formuliert, ist fraglich, ob es zwei vergleichbare regionale Märkte gibt, die sich ohne Mindestlohn gleich entwickelt hätten, aber von einer unterschiedlichen Eingriffsintensität betroffen sind. Eine Antwort auf diese grundlegende Frage bleiben die Studien schuldig.

Die von Dolton et al. (2012a) kürzlich entwickelte Inkrementelle Differenz-von-Differenzen Methode (kurz: IDvD) baut auf diesen regionalen Panelansätzen unter Verwendung regional variierender Eingriffsintensitäten auf. Ähnlich wie Card 1992, Neumark und Wascher 1992 und Stewart 2002 nutzen die Autoren die regionale Variation der Mindestlohn Betroffenheit, die durch regional variierende Löhne zustande kommt, für die Identifikation der Mindestlohneffekte. Insbesondere zielt die IDvD-Methodik darauf ab, die zeitlichen Veränderungen der regionalen Beschäftigung mit Veränderungen der regionalen Betroffenheit zu erklären. Neu an der Methodik ist, dass die Effekte der jährlichen inkrementellen (=stufenweisen) Erhöhungen des Mindestlohnlevels explizit geschätzt werden. Dies ist deshalb interessant, da sich der Gesamteffekt zwar über alle Jahre nicht signifikant von Null unterscheiden kann, dies jedoch nicht zwangsläufig heißen muss, dass die Veränderungen der individuellen Beschäftigungssituation null ist. Der langfristige Gesamteffekt einer einfachen DvD Schätzung ist nicht gleich der Summe aller jährlichen Koeffizienten des IDvD (siehe Dolton et al. 2012a, S. 86). Der Grund hierfür ist, dass der inkrementelle Effekt für sich ändernde Treatmentgruppen im Durchschnitt geschätzt wird. Bei heterogenen Treatmenteffekten kann dies mit einer dynamischen Selektion im Zeitverlauf einhergehen.

Zum anderen unterscheidet sich IDvD von seinen Vorgängerstudien in der identifizierenden Variation. Während die bisherigen Studien den Effekt der

regionalen Betroffenheitsquote auf die regionale Beschäftigung analysieren, schätzt IDvD den Effekt der regional variierenden Musters der inkrementellen Veränderungen in der Betroffenheit über die Zeit. Konkret werden zur Umsetzung der Methodik Indikatoren der regionalen Beschäftigung im Rahmen eines Panelansatzes auf den Betroffenheitsindikator P_{jt} für Region j und Periode t regressiert

$$E_{jt} = \gamma_0 + J_j + \sum_{\tau=1}^T \gamma_{\tau} Y_{\tau} + \theta_0 P_{jt} + \sum_{\tau=l}^T \theta_{\tau}^{IDvD} Y_{\tau} P_{jt} + \delta X_{jt} + \varepsilon_{jt},$$

wobei J_j regionale fixe Effekte und Y_{τ} Zeiteffekte (Jahresdummies) abbilden, während X_{jt} weitere Kovariate bezeichnet. l bezeichne die Periode der Mindestlohneinführung. Die Koeffizienten der Interaktionsterme θ_{τ}^{IDvD} – die Interaktion zwischen der Betroffenheitsquote und den Jahresdummies – bilden dabei die jährlichen IDvD-Koeffizienten. Sie geben den durchschnittlichen Effekt einer jeweiligen Mindestlohnerhöhung relativ zur Periode vor der Einführung ($t < l$) an. Für strukturelle Unterschiede zwischen den Regionen wird in der Schätzung somit unter anderem durch den auch für die Zeit vor Einführung des Mindestlohns definierten Betroffenheitsindikator kontrolliert, dessen Effekt mit θ_0 geschätzt wird. Zusätzlich fangen regionale fixe Effekte regionalspezifische Einflüsse, während die Zeiteffekte alle Faktoren aufgreifen, die über die Zeit, nicht aber über Regionen variieren.

Die identifizierende Variation ist somit das regional variierende Muster der Veränderungen der Betroffenheit im Vergleich zu einer Vorperiode. Der sogenannte Markup, also der Anstieg in der Betroffenheit vom Mindestlohn von einem zum nächsten Jahr wird somit konditional auf Kovariate und den fortlaufenden Betroffenheitsindikator als exogen angenommen. Unter der Annahme, dass sich die regionalen Märkte nicht gegenseitig beeinflussen (SUTVA) und sich die Märkte ohne Mindestlohnerhöhungen ähnlich entwickelt hätten (Common Trends Annahme), lassen sich somit die kausalen Wirkungen der inkrementellen Erhöhungen des ML gegenüber der Vorperiode bestimmen (vgl. Dolton et al. 2012a).

Das Problem räumlicher Korrelation kann durch die Verwendung von Arbeitsmarktregionen (berechnet beispielsweise auf Basis von Pendlerströmen) minimiert werden, da diese relativ abgeschlossene Märkte darstellen, wo Beschäftigte wohnen und arbeiten. Dennoch ist anzunehmen, dass auch zwischen solchen Arbeitsmärkten Interaktionen bestehen, so dass die wirtschaft-

liche Situation in einem Arbeitsmarkt nicht als unabhängig von der Situation in benachbarten Arbeitsmärkten angesehen werden kann (Verletzung der SUT-VA). Eine Möglichkeit der Berücksichtigung solcher Nachbarschaftseffekte bzw. räumlicher Spillover ist die Modellierung räumlich geclusteter Standardfehler oder die Berechnung robuster Standardfehler (robust gegenüber Heteroskedastizität, Wooldridge 2010). Um die räumlichen Interdependenzen explizit zu modellieren, müssten räumlich-ökonometrische Analysen durchgeführt werden. Die Studie von Dolton et al. (2012b) setzt genau hier an und erweitert die IDvD Methodik um eine räumliche Komponente, indem sie mithilfe eines *Spatial Autoregression Panel Model (SARP)* sowie eines *Spatial Error Panel Model (SEMP)*, Elhorst 2010) die räumliche Dimension der abhängigen und unabhängigen Variablen als auch die räumliche Beziehung des Fehlerterms (unbeobachtete Faktoren) berücksichtigen.

Ein schwerwiegenderes Problem bei der Umsetzung des IDvD-Ansatzes ist jedoch die Annahme, dass der fortlaufende „Bite“, also die Eingriffsintensität, exogen ist. Während die räumliche Variation in der Betroffenheit vom Mindestlohn im Fall einer erstmaligen Einführung des Mindestlohns konditional auf strukturelle Unterschiede in den regionalen Märkten eventuell als exogen betrachtet werden kann, sind die folgenden Erhöhungen des „Bite“ hingegen auch eine mögliche Reaktion auf die Betroffenheit im Mindestlohn und somit vermutlich endogen. Die IDvD-Koeffizienten über eine Mindestlohneinführung hinaus sind unseres Erachtens somit kaum kausal interpretierbar (siehe auch die diesbezüglichen Ausführungen in Kapitel 2.1 im Hinblick auf IDvD-Ergebnisse in der vorliegenden Evaluation für das Bauhauptgewerbe).

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne:

Grundsätzlich ist die Methodik für die Evaluierung von Mindestlöhnen in Deutschland anwendbar: Zum einen variiert die Betroffenheit vom Mindestlohn regional stark (siehe bspw. Frings et al. 2013), zum anderen liegen vor allem in den Branchen des Bauhaupt- und Baunebengewerbes lange Zeitreihen vor, eine wichtige Voraussetzung für den Panelansatz. Die Methodik wurde bisher lediglich in Großbritannien (Dolton et al. 2010, 2012a, 2012b) und im Rahmen der Mindestlohnevaluation des Bauhauptgewerbes (IAB/RWI/ISG 2011) angewendet.

Als Vorteil der Methode lässt sich herausheben, dass sie unter starken Annahmen die Abschätzung jeder inkrementellen (stufenweisen) Erhöhung des Mindestlohns erlaubt. Dies ist insbesondere für Deutschland, mit höchst unterschiedlichen und branchenspezifischen Mindestlohnverläufen, interessant. Mit der Methodik könnten neue Einblicke in den Verlauf der Mindestlohnwirkung gewonnen werden. Weiterhin nimmt der Schätzansatz auch langfristige Effekte in den Blick und differenziert zeitinvariante regionale Heterogenitäten heraus, was als Vorteil gegenüber herkömmlichen Methoden gewertet werden kann.

Allerdings weist die Methodik auch klare Nachteile auf. So ist insbesondere die Exogenitätsannahme der Betroffenheitsindikatoren als kritisch einzustufen, da nicht davon ausgegangen werden kann, dass die Betroffenheit einer Region unabhängig von regionalen Strukturmerkmalen ist. So ist beispielsweise nicht auszuschließen, dass Hochlohnregionen mit einer niedrigen Betroffenheit bestimmte endogene regionale Merkmale aufweisen, die wiederum die Ergebnisvariable positiv beeinflussen können. Dies macht es u.a. auch unmöglich, kausale *Lohn*effekte messen zu können, da eine Regression von Lohnungleichheitsparametern auf regionale Betroffenheitsindikatoren (die vom Lohnniveau abhängen) offensichtlich von Endogenitäten geprägt wären. Schließlich ist nicht ganz klar, wie die IDvD Koeffizienten angesichts einer möglichen Vermischung von verschiedenen Effekten zu interpretieren sind. So spiegelt der Koeffizient einer Mindestloohnerhöhungsstufe möglicherweise auch den verzögerten Effekt der vorherigen Erhöhung oder den vorausgehenden (Antizipations-) Effekt der nächsten Erhöhungsstufe wider.

Insgesamt ist die IDvD Methodik als interessante Neuentwicklung zu bewerten, sollte aber vor dem Hintergrund der genannten Endogenitätsprobleme kritisch betrachtet werden.

3.1.4 Synthetische Kontrollgruppen

Der Ansatz synthetischer Kontrollgruppen wurde von Abadie und Gardeazabal (2003) und Abadie et al. (2010) entwickelt, um Interventionen zu evaluieren, die auf aggregierter Ebene implementiert werden (z.B. auf Länder-, Regionen-, oder Branchenebene). Voraussetzung für die Verwendung des Ansatzes ist, dass eine Einheit auf aggregierter Ebene (z.B. eine Region) von der Intervention betroffen ist (Treatmenteinheit), während es die anderen Einheiten (bzw.

Regionen) nicht sind. Ferner wird vorausgesetzt, dass alle Einheiten für einen gewissen Zeitraum vor und nach der Intervention beobachtet sind.

Um den Effekt der Intervention auf eine interessierende Ergebnisvariable zu bestimmen, ist die Frage zu klären, wie sich die Ergebnisvariable in der von der Intervention betroffenen Treatmenteinheit ohne die Intervention entwickelt hätte. Zur Beantwortung dieser Frage ist eine geeignete Kontrollgruppe aus den nicht betroffenen Einheiten (Kontrolleinheiten) zu bilden. Die Idee des Ansatzes synthetischer Kontrollgruppen ist es, alle Kontrolleinheiten mit einem Gewicht in die Kontrollgruppenbildung einzubeziehen. Jeder Kontrolleinheit wird ein Gewicht zwischen 0 und 1 zugeordnet, wobei sich die Gewichte über alle Kontrolleinheiten zu 1 addieren. Die einzelnen Kontrolleinheiten werden auf diese Weise zu einer synthetischen Kontrollgruppe zusammengeführt.

Zur Bestimmung der Gewichte wird ausschließlich auf die zugrundeliegenden Daten zurückgegriffen, um Subjektivität durch den Evaluator bzw. die Evaluatorin auszuschließen. Die Gewichte werden so gewählt, dass sich die Treatmenteinheit und die durch die Gewichte entstehende synthetische Kontrollgruppe im Hinblick auf die Ergebnisvariable oder deren Determinanten für den Zeitraum vor der Intervention möglichst ähneln. Die Auswahl der Gewichte ist folglich mit einem Matching-Schätzer vergleichbar, bei dem statistische Zwillinge erzeugt werden. Die Literatur liefert keinen klaren Hinweis darauf, ob zur Bestimmung der Gewichte auf die Ergebnisvariable vor der Intervention oder deren Determinanten gematcht werden sollte. Für den Fall, dass ein langer Beobachtungszeitraum vor der Intervention zur Verfügung steht, empfehlen Abadie et al. (2010) auf die Ergebnisvariable zu matchen, da damit potenziell auch für unbeobachtete Determinanten der Ergebnisvariablen kontrolliert werden kann. Es bietet sich ferner an, auf die Ergebnisvariable in der Vergangenheit zu matchen, wenn man sicher ist, dass Antizipationseffekte (Ashenfelters Tal) ausgeschlossen werden können und davon auszugehen ist, dass die Ergebnisse in der Vergangenheit eine gute Prognose der zukünftigen Werte der Ergebnisvariablen erlauben.

Anders als beim DvD-Matching-Ansatz stellt der synthetische Kontrollgruppenansatz eine Vergleichbarkeit für die Zeit vor der Mindestlohneinführung auf Branchenebene her, so dass die Common Trends Annahme somit eher

erfüllt sein sollte als bei einem Ansatz, der die Vergleichbarkeit nur auf Individualebene gewährleistet.

Die ermittelten Gewichte bzw. die synthetische Kontrollgruppe wird anschließend genutzt, um für die Zeit nach der Intervention die hypothetische Entwicklung der Ergebnisvariablen für die Treatmenteinheit aufzuzeigen, die sich ohne die Intervention ergeben hätte. Alle Unterschiede in der Ergebnisvariablen nach der Intervention zwischen Treatmenteinheit und synthetischer Kontrollgruppe sind auf die Intervention zurückzuführen und stellen den Treatmenteffekt der Intervention dar.

Bei der Ermittlung der Gewichte ist zu beachten, dass bei einer hinreichend großen Anzahl von Kontrolleinheiten relativ zur Anzahl der anzupassenden Ergebnis- bzw. Vorhersagevariablen das resultierende Gleichungssystem unteridentifiziert ist. In diesem Fall kann es mehrere Lösungen für die optimalen Gewichte geben. Als Test der Eignung der Gewichte bietet es sich daher an, diese zunächst mit einer Teilstichprobe der Prä-Interventionsperiode zu ermitteln und die Qualität der Gewichte anhand einer „Out-of-Sample“-Vorhersage innerhalb der Prä-Interventionsperiode zu überprüfen. Falls ein ausreichend langer Beobachtungszeitraum zur Verfügung steht, sollten die Gewichte für mehrere Teilintervalle in der Prä-Interventionsperiode berechnet und über alle Intervalle hinweg verglichen werden. Je stabiler die Gewichte über die Zeit sind, desto verlässlicher können sie in die Analyse einbezogen werden.

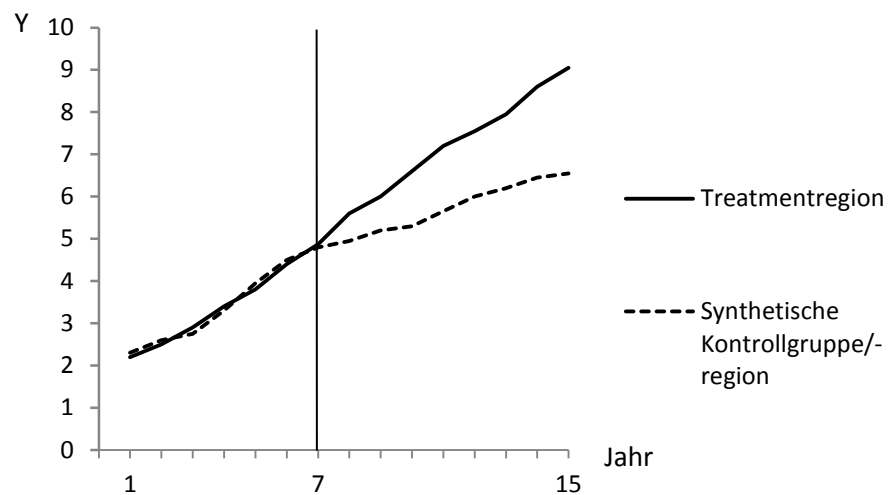
Zum besseren Verständnis des Ansatzes synthetischer Kontrollgruppen ist in Abbildung 1 ein Beispielfall illustriert. In diesem Beispiel sind 40 Regionen über einen Zeitraum von 15 Jahren beobachtet. Eine Region unterliegt nach 7 Jahren einer politischen Intervention. Alle anderen Regionen sind von dieser Intervention nicht betroffen. Um den Effekt der Intervention auf die Ergebnisvariable Y zu bestimmen, werden die 39 nicht betroffenen Regionen zu einer synthetischen Kontrollgruppe bzw. Kontrollregion zusammengefasst. Dazu erhält jede nicht betroffene Region ein Gewicht zwischen 0 und 1.¹² Wie die

¹² In den Anwendungsfällen des Ansatzes synthetischer Kontrollgruppen in der Literatur ist es nicht unüblich, dass nur einzelne Kontrolleinheiten ein positives Gewicht erhalten und

Abbildung verdeutlicht, erfolgt die Gewichtung so, dass möglichst kein Unterschied in der Ergebnisvariable der Treatmentregion und der synthetischen Kontrollregion für die Zeit vor der Intervention in Jahr 7 besteht.

Werden die Gewichte auf die Zeit nach der Intervention übertragen, zeigt sich, dass die Ergebnisvariable Y in der Treatmentregion größer ist als in der synthetischen Kontrollregion. Dieser Unterschied zwischen Treatment- und synthetischer Kontrollregion ist auf die Intervention zurückzuführen. Die Intervention hat also einen positiven Effekt, da sich die Ergebnisvariable erhöht.

Abbildung 1: Verlauf von Y in Treatment- und synthetischer Kontrollregion



Die Signifikanz des auf diese Weise ermittelten Effektes lässt sich mit Hilfe so genannter Placebotests ermitteln. Hierzu wird für jede der betrachteten Kontrolleinheiten der Ansatz synthetischer Kontrollgruppen mit dieser Kontrolleinheit als fiktiver Treatmenteinheit durchgeführt. Alle übrigen Einheiten, einschließlich der eigentlichen Treatmenteinheit, werden als Kontrolleinheiten herangezogen. Im oben skizzierten Beispielfall würden also 39 Placeboeffekte

die restlichen Kontrolleinheiten mit einem Gewicht von 0 aus der Analyse ausscheiden (siehe z.B. Abadie und Gardeazabal 2003 sowie Abadie et al. 2010).

ermittelt. Die so ermittelten 39 Placeboeffekte werden nun mit dem Effekt für die tatsächliche Treatmenteinheit verglichen. Da den Placebotests kein reales Treatment zugrunde liegt, sollten die Placeboeffekte gegen Null tendieren und kleiner ausfallen als der Effekt für die tatsächliche Treatmenteinheit. Wenn die überwiegende Mehrheit (z.B. 95%) der Placeboeffekte kleiner ist als der Effekt für die tatsächliche Treatmenteinheit, so kann letzterer als statistisch signifikant angenommen werden.

Der Ansatz synthetischer Kontrollgruppen wurde in der Mindestlohnforschung bisher nur sehr selten eingesetzt. Lediglich in drei Studien für die USA (Sabia et al. 2012, Neumark et al. 2013 und Allegretto et al. 2013) wird der Ansatz oder Teile davon verwendet. Die Anwendung erfolgt allerdings zumeist nur im Rahmen von Sensitivitätsanalysen.

Sabia et al. (2012) untersuchen eine Mindestloohnerhöhung im Bundesstaat New York von 5,15\$ auf 6,75\$ in den Jahren von 2004 bis 2006. Ihre Hauptuntersuchung basiert auf einem DvD-Ansatz, wobei als Kontrollgruppe Personen aus benachbarten Bundesstaaten (Pennsylvania, Ohio und New Hampshire) herangezogen werden. Die Ergebnisse zeigen, dass der Mindestlohnanstieg in New York die Beschäftigungswahrscheinlichkeit Geringqualifizierter im Alter von 16 bis 29 Jahren um 20 Prozent verringert hat.

Um zu analysieren, ob dieses Ergebnis durch die Auswahl der Kontrollgruppe beeinflusst ist, aggregieren die Autoren ihre individuellen Daten auf Bundesstaatebene und bilden eine synthetische Kontrollgruppe für den Bundesstaat New York aus allen anderen Bundesstaaten der USA, die im Zeitraum von 2002 bis 2006 einen Mindestlohn von 5,15\$ hatten.¹³ Allerdings gelingt es ihnen nicht, die synthetische Kontrollgruppe präzise an die Beschäftigungssituation in New York vor der Erhöhung des Mindestlohns anzupassen. Daher ist es ihnen nicht möglich, lediglich die Beschäftigungssituation nach der Mindestloohnerhöhung zwischen New York und der synthetischen Kontrollgruppe zu vergleichen, sondern sie müssen einen DvD-Ansatz einsetzen, um für die Un-

¹³ In die synthetische Kontrollgruppe gehen Pennsylvania, Ohio, Colorado, Maryland, Michigan, Nevada und Virginia ein. New Hampshire, das ursprünglich als Kontrollgruppe für den DvD-Ansatz ausgewählt wurde, erhält ein Gewicht von 0.

terschiede zwischen beiden Gruppen vor Erhöhung des Mindestlohns zu kontrollieren. Die Ergebnisse dieses DvD-Ansatzes bestätigen die Resultate der Hauptuntersuchung.

Neumark et al. (2013) diskutieren in ihrer Studie die Ergebnisse von Allegretto et al. (2011) und Dube et al. (2010). Während Allegretto et al. (2011) die Effekte von Mindestlohnerhöhungen in den USA auf die Löhne und Beschäftigungschancen von Jugendlichen für den Zeitraum von 1990 bis 2009 untersuchen, analysieren Dube et al. (2010) die Auswirkungen von Mindestlohnerhöhungen auf Beschäftigte im Niedriglohnbereich der USA von 1990 bis 2006. Beide Studien, die keine Beschäftigungswirkung der Erhöhung des Mindestlohns auf individueller Ebene feststellen können, nutzen in ihrer Untersuchung Veränderungen des Mindestlohns auf Bundesstaatebene und ziehen als Kontrollgruppe Personen aus geografisch benachbarten Bundesstaaten heran, die zudem in derselben US Census Division (Allegretto et al. 2011) bzw. in aneinander angrenzenden Landkreisen wohnen (Dube et al. 2010). Neumark et al. (2013) kritisieren diese Kontrollgruppen-Auswahl. Sie zeigen mit einem auf Bundesstaat-Ebene aggregierten Datensatz, dass bei Verwendung des Ansatzes synthetischer Kontrollgruppen die von Allegretto et al. (2011) und Dube et al. (2010) gewählten Bundesstaaten nur ein geringes Gewicht erhalten würden, sie also als Kontrollgruppe ungeeignet sind. Allerdings führen Neumark et al. (2013) den Ansatz synthetischer Kontrollgruppen nicht vollständig aus. Sie ermitteln nur Gewichte für die einzelnen Kontroll-Bundesstaaten, analysieren aber keine Unterschiede zwischen den Bundesstaaten mit Mindestlohnerhöhung und der jeweils zugehörigen synthetischen Kontrollgruppe.

Um der Kritik von Neumark et al. (2013) zu begegnen, führen Allegretto et al. (2013) eine weitere Analyse durch. In dieser Analyse kritisieren die Autoren wiederum die Vorgehensweise von Neumark et al. (2013) bei der Bestimmung der Gewichte im Rahmen des Ansatzes synthetischer Kontrollgruppen. Einer ihrer Hauptkritikpunkte ist, dass Neumark et al. (2013) nur einen begrenzten Beobachtungszeitraum vor den untersuchten Mindestlohnerhöhungen zur Verfügung haben (4 Quartale) und die Matchqualität von Treatmentgruppe und synthetischer Kontrollgruppe dadurch negativ beeinträchtigt wird. Des Weiteren sind die Variablen, die in das Matching einbezogen werden, aus Sicht von Allegretto et al. (2013) durch frühere Mindestlohnerhöhungen beeinflusst, so dass die von Neumark et al. (2013) ermittelten Gewichte keine validen

Schlussfolgerungen darüber zulassen, ob die von Allegretto et al. (2011) sowie von Dube et al. (2010) verwendeten Kontrollgruppen geeignet sind oder nicht.

Zwar führen Allegretto et al. (2013) keine eigenen Analysen zur Wirkung von Mindestloohnerhöhungen auf Basis des Ansatzes synthetischer Kontrollgruppen durch, aber sie reproduzieren Ergebnisse aus einer unveröffentlichten Studie von Dube und Zipperer (2013). Dube und Zipperer (2013) analysieren die Wirkung von 19 Mindestloohnerhöhungen in den USA auf die Beschäftigungschancen von Jugendlichen mittels des Ansatzes synthetischer Kontrollgruppen, wobei sie den Neumark et al. (2013) vorgeworfenen Problemen mit einem längeren Beobachtungszeitraum und besser geeigneten Matchvariablen aus dem Wege gehen. Die Ergebnisse von Dube und Zipperer (2013) zeigen, dass die untersuchten Mindestloohnerhöhungen im Durchschnitt keinen Einfluss auf die Beschäftigungswahrscheinlichkeit der Jugendlichen hatten. Somit bestätigen sie die Ergebnisse von Allegretto et al. (2011) sowie Dube et al. (2010).

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne: Auch wenn der Ansatz synthetischer Kontrollgruppen bisher nur selten in der Mindestlohnforschung eingesetzt wurde, hat der Ansatz für die Evaluation von Branchenmindestlöhnen großes Potenzial. Da in Deutschland Mindestlöhne auf Branchenebene eingeführt wurden, ist die Voraussetzung einer Intervention auf aggregierter Ebene gegeben. Prinzipiell ließe sich somit für eine Branche, die einen Mindestlohn eingeführt hat, eine synthetische Kontrollgruppe aus anderen Branchen bilden. Einflüsse des Mindestlohns auf Variablen, die auf Branchenebene gemessen werden (wie z.B. das Beschäftigungswachstum in einer Branche), könnten dann bestimmt werden. Der Einfluss des Mindestlohns auf individuelle Ergebnisvariablen ließe sich jedoch nicht ermitteln.

Insgesamt weist der synthetische Kontrollgruppenansatz den Vorteil auf, dass zur Identifizierung einer geeigneten Kontrollgruppe die Variation über alle Kontrolleinheiten genutzt wird, anstatt wie in den bisherigen Evaluationen alle Kontrolleinheiten im Rahmen der Prüfung ihrer Validität gleich zu gewichten. Eine weitere Chance des Ansatzes besteht darin, dass – sofern zur Konstruktion der synthetischen Kontrollgruppe externe Vergleichsbranchen zum Einsatz kommen – die Vorteile des Kontrollgruppenansatzes außerhalb der betroffenen Branche genutzt werden, seine Nachteile aber durch das Matching der

Ergebnisvariablen auf den entsprechenden Verlauf in der Treatmentbranche beseitigt werden können.

Bei der konkreten Anwendung des Ansatzes stellt sich die Frage, welche Vergleichsbranchen in den so genannten „Donorpool“ zur Konstruktion der synthetischen Kontrollgruppe mit einbezogen werden sollen. In diesem Kontext weisen Abadie et al. (2012) darauf hin, dass „because comparison units are meant to approximate the counterfactual of the case of interest without the intervention, it is important to restrict the donor pool to units with outcomes that are thought to be driven by the same structural process as the unit representing the case of interest and that were not subject to structural shocks to the outcome variable during the sample period of the study“. Dies legt nahe für die Vergleichsbranchen im Donorpool Branchen ohne Mindestlöhne, aber mit möglichst ähnlichen Strukturen auf den zugrundeliegenden Güter- und Arbeitsmärkten heranzuziehen.¹⁴ Zudem sollten die Vergleichsbranchen nicht zur Zeit der Intervention in der Treatmentbranche oder später während des Beobachtungszeitraums strukturellen Veränderungen ausgesetzt sein. Sollte diese Annahme für die herangezogenen Vergleichsbranchen wider Erwarten verletzt sein, kann dies jedoch durch die oben angesprochenen Placebotests aufgedeckt werden. Als Nachteil einer ex-ante Beschränkung der Vergleichsbranchen im Donorpool ist jedoch darauf hinzuweisen, dass eine zu starke Restringierung unter Umständen die Konstruktion einer passenden synthetischen Kontrollgruppe behindern kann. Eine ausführlichere Diskussion möglicher Umsetzungsprobleme unter Berücksichtigung der Branchenspezifika erfolgt in Kapitel 5.

3.2 Alternative Verfahren zum DvD-Ansatz

Neben dem DvD-Ansatz und dessen Varianten gibt es weitere ex-post Ansätze zur Identifikation von kausalen Effekten von Politikinterventionen. In den nächsten Unterkapiteln wird daher näher auf die Regressions-

¹⁴ Dieser Ansatz wird implizit von Chung et al. (2013) verfolgt, die bei der Evaluation einer branchenspezifischen Handelspolitik in den USA zur Konstruktion der synthetischen Kontrollgruppe Branchen aus der benachbarten Wirtschaftszweigklassifikation heranziehen.

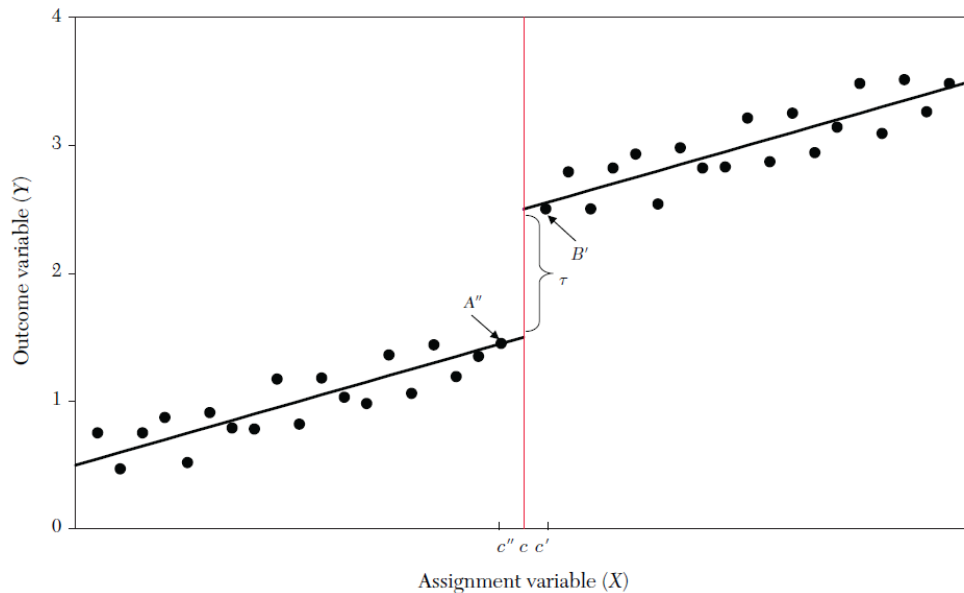
Diskontinuitätsanalyse, den Instrumentenvariablenansatz sowie den Kontrollfunktionsansatz eingegangen.

3.2.1 Discontinuity-Designs

Die grundlegende Idee der Regressions-Diskontinuitätsanalyse (RD-Analyse) besteht darin, dass die Zuteilung in Kontroll- bzw. Treatmentgruppe durch einen Schwellenwert in einer beobachteten kontinuierlichen Kontrollvariable X bestimmt wird, der als exogen angesehen werden kann. Die Treatmentgruppe (Kontrollgruppe) rekrutiert sich aus Beobachtungen mit Werten der betreffenden Variablen beispielsweise oberhalb (unterhalb) des Schwellenwertes. Als häufig genutzte Beispiele für Variablen, die zu einer exogenen Zuordnung in eine Treatment- und Kontrollgruppe anhand eines Schwellenwertes führen, sind insbesondere Alter, Geburtsjahr, Testergebnisse, der Stimmenanteil bei Wahlen sowie eine regionale (meist administrative) Grenze. Ähnlich wie beim DvD-Ansatz wird somit eine quasi-experimentelle Situation ausgenutzt. Anders als der DvD-Ansatz erfordert der RD-Ansatz nicht zwingend Längsschnittdaten vor und nach der Politikintervention und kann auf Basis von Querschnittdaten angewendet werden. Dies ist möglich, da Treatment- und Kontrollgruppe sich durch die exogene Zuordnung annahmegemäß nicht in unbeobachteten Faktoren unterscheiden (siehe unten), so dass eine Beseitigung unbeobachteter Unterschiede durch Differenzenbildung anders als beim DvD-Ansatz nicht erforderlich ist.

Bei der konkreten Umsetzung eines RD-Designs sind grundsätzlich zwei verschiedene Typen zu unterscheiden. Beim scharfen RD-Design („sharp design“) ist die Unterteilung in Treatment- und Kontrollgruppe eine deterministische Funktion der zugrundeliegenden Kontrollvariable. Dies wird in Abbildung 2 veranschaulicht, in der die Variable X , die für die zu erklärende Variable Y am Punkt c eine Diskontinuität in Höhe von τ hervorruft, eindeutig in eine Kontroll- (Werte von X links der Diskontinuität) und in eine Treatmentgruppe (Werte von X rechts der Diskontinuität) aufgeteilt werden kann. Bei der unscharfen RD-Analyse („fuzzy design“) dagegen bestimmt die Kontrollvariable die Einteilung in Treatment- und Kontrollgruppe nicht deterministisch, beeinflusst aber ihre Wahrscheinlichkeit oder ihren Erwartungswert.

Abbildung 2: Beispiel für ein scharfes RD-Design



Quelle: Lee und Lemieux (2010)

Hahn et al. (2001) zeigen für das RD-Design, dass die Identifikation eines kausalen durchschnittlichen (lokalen) Treatmenteffektes nur einige wenige Annahmen benötigt. Zum einen muss eine diskontinuierliche Beziehung zwischen der exogen variierenden Variable X sowie dem Treatment-Status vorliegen. Zum anderen muss angenommen werden, dass sich die beiden Gruppen knapp oberhalb und knapp unterhalb des durch die Diskontinuität bestimmten Schwellenwerts in Bezug auf die Verteilung aller weiteren beobachteten und unbeobachteten Variablen (außer X) gleichen. Ob diese Annahme zutrifft, kann mittels sogenannter Balance-Tests deskriptiv leicht überprüft werden, indem die Ausprägungen verschiedener Kontrollvariablen zwischen beiden Gruppen auf systematische Unterschiede untersucht werden. Die Gültigkeit der Annahmen führt dazu, dass jede Diskontinuität in der betrachteten Ergebnisvariable am Schwellenwert auf den Treatmentstatus zurückgeführt werden kann. Dies setzt jedoch drei zusätzliche Annahmen voraus.

Zum einen darf es nicht zu einer nachträglichen Selektion von Beobachtungen in das Treatment kommt. Während dies im Fall beispielsweise eines Alters-Schwellenwertes eine unkritische Annahme darstellt, da das Alter nicht durch

das Individuum beeinflusst werden kann, ist diese Annahme im Fall einer regionalen Grenze möglicherweise problematisch, wenn Individuen oder Betriebe ihren Standort und damit die Zugehörigkeit zum Treatment selbst bestimmen können.

Zweitens darf es keine Antizipationseffekte geben, wenn Individuen z.B. im Falle einer Altersgrenze automatisch von der Kontroll- in die Treatmentgruppe wechseln. Kommt es in der Kontrollgruppe somit bereits zu Verhaltensanpassungen bevor die Altersgrenze zum Treatment überschritten wird, ist mit einer Unterschätzung des Treatment-Effektes zu rechnen.

Drittens darf es zu keinen Wechselwirkungen zwischen Kontroll- und Treatmentgruppe kommen. Ähnlich wie beim DvD-Design ist jedoch auch beim RD-Design mit möglichen Wechselwirkungen in Form von Substitutions- und Skaleneffekten zu rechnen.

Darüber hinaus besteht eine Schwierigkeit bei der Umsetzung eines RD-Designs in der Wahl der richtigen funktionalen Form (Lee und Lemieux 2010). Da der RD-Ansatz den kausalen Effekt des Treatments schätzt, indem der Zusammenhang zwischen der Ergebnisvariable und dem Abstand vom Schwellenwert in der zuordnenden Variable X modelliert wird, muss eine funktionale Form des Zusammenhangs angenommen werden. Dabei ist jedoch unklar, welche funktionale Form tatsächlich zugrunde liegt. Folgt das „wahre“, zugrundeliegende Modell beispielsweise keinem linearen Zusammenhang, wie möglicherweise angenommen, so wäre eine Schätzung im Allgemeinen verzerrt und nicht erwartungstreu. Mögliche Abhilfen hierfür sind das Einfügen beispielsweise quadratischer Terme, um mögliche Nichtlinearitäten zu erfassen. Die Ergebnisse von bspw. Dickens et al. (2011) sind relativ robust gegenüber verschiedenen funktionalen Formen der Altersvariable.

Unabhängig davon, welche funktionale Form angenommen wird, wird dem Modell durch die Parametrisierung eine möglicherweise ungerechtfertigte Restriktion auferlegt. Dies lässt sich durch nicht-parametrische Analysen umgehen, siehe z.B. Hahn et al. (2001) und Olssen (2011). Die Idee hierbei ist, dass lokale lineare Regressionen in verschiedenen Intervallen oberhalb und unterhalb des Schwellenwertes geschätzt werden, wobei die Bandbreite festlegt, wie groß das Intervall ist. Bei der Wahl der Bandbreite besteht ein Trade-Off zwischen der Präzision und der Verzerrung der Ergebnisse. Wird die Band-

breite klein gewählt, so dass die Gruppe unterhalb und oberhalb des Schwellenwertes klein ist, ist eine gute Präzision der Ergebnisse möglicherweise durch zu kleine Beobachtungszahlen nicht gewährleistet. Ist die Bandbreite groß, ist die Präzision hoch, dafür ist eine Verzerrung der Ergebnisse möglich, da Personen in die Analyse einfließen, die weiter weg vom Schwellenwert liegen.

RD-Ansätze wurden in der Mindestlohnforschung bereits mehrfach angewendet. Ein Beispiel für den RD-Ansatz zur Analyse der kausalen Effekte von Mindestlöhnen stammt von Dickens et al. (2011). Die Autoren nutzen aus, dass die Höhe des Mindestlohns in Großbritannien auch vom Alter der Beschäftigten abhängt. So liegt der Mindestlohn für erwachsene Personen ab 22 Jahren um etwa 20% oberhalb des Mindestlohns für Beschäftigte unterhalb von 22 Jahren. Dickens et al. (2011) nutzen diese Diskontinuität in der Höhe des Mindestlohns und vergleichen in dieser quasi-experimentellen Situation Beschäftigte, die bereits 22 Jahre alt sind, mit denen, die noch 21 Jahre alt sind. Die Zielgruppe sind also solche Individuen, die nicht mehr als 12 Monate von der Diskontinuität entfernt sind. Verwendete Zielgrößen sind sowohl die Beschäftigungs- und Arbeitslosigkeitswahrscheinlichkeit als auch die Wahrscheinlichkeit auf dem Arbeitsmarkt inaktiv zu sein.

Die Autoren nehmen an, dass sich Personen oberhalb des Schwellenwertes von 22 Jahren (Treatmentgruppe) in ihren beobachteten und unbeobachteten Charakteristika nicht von den Beschäftigten unterhalb des Schwellenwertes (Kontrollgruppe) unterscheiden. Folglich ist die Zuteilung der Personen in die beiden Gruppen als zufällig anzusehen, solange allein das Alter, welches nicht von den Beschäftigten beeinflusst werden kann, diese Zuteilung bestimmt. Unter dieser identifizierenden Annahme können Unterschiede in den Ergebnisvariablen auf die Unterschiede im Mindestlohn an der Altersschwelle zurückgeführt werden. Die Ergebnisse zeigen, dass die Wahrscheinlichkeit beschäftigt zu sein, um 2-4 Prozent steigt, wenn Personen 22 Jahre alt werden. Zudem sinkt besonders für Männer die Wahrscheinlichkeit arbeitslos zu sein, während für Frauen tendenziell die Inaktivität am Arbeitsplatz mit dem Erreichen des 23. Lebensjahres abnimmt.

Im Fall der von Dickens et al. genutzten Altersgrenze wechselt jeder Beschäftigte zu einem gewissen Zeitpunkt in der Zukunft in die Treatmentgruppe, so

dass es zu den oben diskutierten Antizipationseffekten kommen kann. Dickens et al. schätzen solche Effekte jedoch als gering ein, da die hohe Fluktuation bei der Beschäftigung im Niedriglohnsektor, auf den sich die Studie konzentriert, nicht dazu führen sollte, dass junge Beschäftigte ihr Verhalten von der Mindestlohnhöhe abhängig machen.

Olssen (2011) nutzt ebenfalls ein scharfes RD-Design, um mit Hilfe von altersbedingten Mindestloohnerhöhungen in Australien u.a. die Auswirkungen von Mindestlöhnen auf die geleisteten Arbeitsstunden zu identifizieren. Die Höhe des Mindestlohns variiert in Australien dabei nicht nur nach dem Alter, sondern auch nach Regionen und Branchen. Konkret gilt ab dem Alter von 16 Jahren in Australien ein Mindestlohn, der für 16-Jährige die Hälfte des Mindestlohns beträgt, welchen Erwachsene ab 21 Jahren erhalten. Mit jedem weiteren Lebensjahr steigt der Mindestlohn bis zum 21. Geburtstag stufenweise um 10% an. Olssen (2011) nutzt die Diskontinuitäten in der Höhe des Mindestlohns um den 19., 20. und 21. Geburtstag aus und vergleicht u.a. die geleisteten Arbeitsstunden von Personen, die ein halbes Jahr unterhalb bzw. oberhalb des jeweiligen Schwellenwertes liegen. Seine Ergebnisse zeigen für alle drei Schwellenwerte, dass ein Anstieg des Mindestlohns um 10% die geleisteten Arbeitsstunden nicht verändert.¹⁵

Neben Altersgrenzen wurde in der Mindestlohnliteratur auch eine regionale Variation der Mindestlohnhöhen für RD-Ansätze genutzt. In diesem Fall ist die kontinuierliche Variable X , die zu einer Diskontinuität im Treatment-Status führt, durch die Distanz der Beobachtungseinheiten zur jeweiligen die Region bestimmenden Grenze gegeben. So repliziert Moore (2008) beispielweise die DvD-Studie von Card und Krueger (1994), berücksichtigt aber zusätzlich die

¹⁵ DiNardo und Lee (2004) wenden die RD-Analyse auf Betriebsebene an, um den Einfluss der gewerkschaftlichen Organisation u.a. auf Löhne abzuschätzen. Zur Identifizierung des Effektes nutzen sie die Variation beim Abstimmungsergebnis von Gewerkschaftswahlen aus. So vergleichen sie die Entwicklung der Löhne in Betrieben, in denen Gewerkschaften eine Wahl knapp gewonnen haben mit solchen, in denen Gewerkschaften knapp verloren haben. Die Lohneffekte fallen gering aus, was die Autoren dahingehend interpretieren, dass die bei Abstimmungswahlen gerade gewinnenden Gewerkschaften nicht erfolgreich waren, höhere Lohnzuwächse durchzusetzen.

Distanz der betrachteten Fast-Food-Restaurants bis zur Grenze. Dabei wird angenommen, dass ein Restaurant nahe der Landesgrenze einem Restaurant auf der anderen Seite der Grenze mehr ähneln sollte als ein entsprechend von der Grenze weiter entferntes Restaurant. Die Ergebnisse dieser räumlichen RD-Analyse unterscheiden sich jedoch nicht von den Ergebnissen aus Card und Krueger (1994).¹⁶

Ein weiteres Beispiel für die Anwendung eines räumlichen RD-Ansatzes liefert die Studie von Magruder (2013), der den Effekt von regional variierenden Mindestlöhnen in Indonesien untersucht. Magruder (2013) argumentiert, dass die Verwendung eines RD-Ansatzes möglicherweise die Problematik von Politikendogenitäten verringert, die daraus resultieren können, dass die Höhe von Mindestlöhnen endogen durch die jeweilige Arbeitsmarktlage determiniert werden. Im Falle Indonesiens, so seine Argumentation, seien Politikendogenitäten eher auf aggregierter Ebene von Bedeutung und somit für direkt benachbarte Regionen von geringerer Relevanz.

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne: In den meisten Branchen, die in Deutschland einem allgemeinverbindlichen Mindestlohn unterliegen, gibt es oder gab es eine Differenzierung der Mindestlohnhöhe zwischen West- und Ostdeutschland. Die innerdeutsche Grenze könnte somit eine Diskontinuität im Treatment-Status darstellen, die für eine Kausalanalyse mittels eines RD-Designs geeignet ist. Allerdings gibt es verschiedene Gründe, die dafür notwendige Exogenität des Treatment-Status zu bezweifeln. Dies ist insbesondere der Fall, weil sich der zu zahlende Mindestlohn unter Umständen am Ort der Leistungserbringung und nicht am Standort des Betriebes orientiert. In solch einem Fall könnten ostdeutsche Betriebe somit Leistungen in Westdeutschland nur zum höheren Mindestlohniveau erbringen. Sobald davon ausgegangen werden kann, dass Betriebe grenzüber-

¹⁶ Eine weitere auf räumlichen Diskontinuitäten basierende Analyse ist die Studie von Dube et al. (2010) für die USA. Die Autoren nutzen zur Identifikation der Mindestlohneffekten im Rahmen eines DvD-Ansatzes die Variation von Mindestlöhnen in jeweils angrenzenden Landkreisen. Damit geht die Studie über die Untersuchung von Card und Krueger (1994) hinaus, die diesen Vergleich lediglich für Restaurants für zwei Bundesstaaten (Pennsylvania und New Jersey) durchgeführt hatten.

scheidend tätig sind, führt dies dazu, dass es keine wirkliche Diskontinuität im Treatment gibt, die zu Identifikationszwecken genutzt werden könnte. Wie sehr die potenzielle Diskontinuität im Treatment beiderseits der Grenze jedoch nivelliert wird, sollte von Branche zu Branche stark unterschiedlich sein. Im Fall einer stark ortsgebundenen Leistungserbringung, wie z.B. der von Moore (2008) untersuchten Fast-Food-Restaurants, sollte eine Diskontinuität im Treatment dennoch existieren. Bei Gewerben, deren Leistungserbringung sich nach dem Standort des Kunden richten (z.B. Leistungen des Bauhaupt- und Baunebengewerbes) ist hingegen nur mit einer schwach ausgeprägten Diskontinuität zu rechnen. Zudem besteht auch für die Betriebe, die aufgrund einer ortsgebundenen Leistungserbringung an das regional gültige Mindestlohniveau gebunden sind, die Möglichkeit, ihren Standort zu wählen und sich somit potenziell in das Treatment (bzw. in die Kontrollgruppe) zu selektieren. Auch die Annahme der exogenen Standortwahl ist somit zu hinterfragen.

Selbst wenn eine hinreichende Diskontinuität im Treatment an der Grenze existiert, können Wechselwirkungen jedoch eine kausale Abschätzung mittels des RD-Ansatzes erschweren. Zwar ist es einerseits wünschenswert, dass die Grenze einen eigentlich einheitlichen Markt in zwei Regionen mit unterschiedlichen Mindestlohnhöhen zerteilt, da dies die Vergleichbarkeit von Kontroll- und Treatmentgruppe gewährleistet; falls es sich dies- und jenseits der Grenze jedoch tatsächlich um einen hoch integrierten Markt handelt, sind Wechselwirkungen wahrscheinlich. Auf der anderen Seite gewährleistet eine starke Marktintegration entlang der innerdeutschen Grenze, dass wie im Beispiel von Magruder (2013) Politikendogenitäten verringert werden können, da strukturelle Unterschiede in Ost und Westdeutschland, die eine Ursache der unterschiedlichen Mindestlohnhöhen sind, an der Grenze an Bedeutung verlieren. Gerade diese Marktintegration schafft jedoch durch Spillovereffekte entlang der Grenze wahrscheinlich andere Identifikationsprobleme.

Abschließend lässt sich somit festhalten, dass die Anwendung eines RD-Designs zur Abschätzung kausaler Mindestlohnwirkungen in Deutschland in einigen Branchen vermutlich mit großen Identifikationsproblemen verbunden ist. In diesem Fall dürfte der RD-Ansatz kaum eine Verbesserung gegenüber den bereits angewendeten DvD-Ansätzen darstellen. Eine ausführlichere Diskussion der oben angesprochenen Identifikationsprobleme unter Berücksichtigung der Branchenspezifika erfolgt in Kapitel 5. Zudem ist das RD-Design, im

Gegensatz zum DvD-Ansatz, nicht geeignet, den Gesamteffekt einer Maßnahme, wie z.B. die Einführung oder Erhöhung eines Mindestlohns, auf relevante Zielgrößen zu messen. Stattdessen dient eine RD-Analyse dazu, einen lokalen kausalen Effekt an der Schwelle, an der die Diskontinuität auftritt, zu identifizieren. Hinzu kommt, dass dies typischerweise der lokale Effekt des Unterschiedes im Mindestlohn (bzw. der Veränderung des Unterschiedes) auf den beiden Seiten der Schwelle und nicht der Effekt der Einführung des Mindestlohns (bzw. der Erhöhung der Mindestlohns) auf einer Seite der Schwelle wäre.

3.2.2 Ansätze unter Verwendung von Instrumentenvariablen

Während der DvD-Ansatz zwar ermöglicht zeitkonstanten Unterschieden zwischen Treatment- und Kontrollgruppe Rechnung zu tragen, wird eine Korrelation des Treatment-Status mit zeitvariierenden unbeobachteten Einflussfaktoren nicht berücksichtigt. Wie bereits in Abschnitt 2.1.1 ausgeführt, kann es in Bezug auf den Mindestlohn zu einer solchen Korrelation z.B. infolge von einer Politikendogenität kommen, wenn die Einführung eines Mindestlohns in einer Region oder Branche systematisch mit Veränderungen der Arbeitsmarktlage korreliert ist. So ist denkbar, dass Mindestlöhne möglicherweise gerade dann eingeführt werden, wenn eine Region oder Branche durch besonders negative (oder zu erwartende besonders negative) Lohnentwicklungen geprägt ist.

Unter Verwendung einer Instrumentenvariablen für die endogene Mindestlohnhöhe bzw. Betroffenheit durch den Mindestlohn ist es prinzipiell möglich, die Endogenität des Treatment-Status im Hinblick auf zeitspezifische unbeobachtete Faktoren zu berücksichtigen. So genannte Instrumentenvariablen – also Variablen, die eine exogene Variation der endogenen Variablen bedingen – müssen hierbei zwei Annahmen erfüllen. Zum einen sollte die Instrumentenvariable mit der erklärenden Variablen korreliert sein und keine Erklärungskraft in Bezug auf die zu erklärende Variable aufweisen. Diese Annahme ist testbar. Die zweite wichtige Annahme – die fehlende Korrelation der Instrumentenvariablen mit dem Fehlerterm der Ergebnisgleichung – lässt sich hingegen nicht überprüfen.

Falls beide Annahmen zutreffen, lassen sich zwei verschiedene Ansätze wählen, um mittels der Instrumentenvariablen den kausalen Effekt des Treatments abzuschätzen.

Die übliche Verfahrensweise einer IV-Schätzung entspricht einer zweistufigen Kleinst-Quadrate-Schätzung (2SLS). Im ersten Schritt wird der endogene Treatmentstatus auf alle gültigen Instrumente sowie alle exogenen Variablen regressiert. Da die Instrumente als exogen angenommen werden, sollte die aus dieser Schätzung resultierende Vorhersage der endogenen Variablen, des Treatments, nicht mit dem Fehlerterm korrelieren. Im zweiten Schritt wird die interessierende Regression wie gewohnt geschätzt, wobei alle endogenen erklärenden Variablen durch die Näherungswerte aus dem ersten Schritt ersetzt werden. Falls sich der Treatmenteffekt zwischen Beobachtungen unterscheidet, wird mit dem IV-Ansatz nur der Effekt für diejenigen Beobachtungseinheiten, deren Treatmentstatus sich tatsächlich durch die Variation des Instruments ändert, identifiziert. Aufgrund der Heterogenität der Effekte kann hieraus nicht auf die übrigen Beobachtungseinheiten geschlossen werden, so dass der IV-Ansatz nur einen „lokalen“ Effekt ermittelt.

Alternativ kann die Instrumentenvariable in einer Kontrollfunktion verwendet werden, um den Treatment-Status zu modellieren. Ähnlich wie bei dem IV-Ansatz wird bei dem Kontrollfunktions-Ansatz (KF-Ansatz) dem Problem der Korrelation zeitspezifischer Schocks mit dem Treatment-Status Rechnung getragen. Anders als beim IV-Ansatz wird jedoch im Rahmen des KF-Ansatzes der Selektionsprozess in den Treatment-Status explizit durch eine Selektionsgleichung modelliert, so dass der KF-Ansatz Aufschluss über zugrundeliegende strukturelle Determinanten des Selektionsprozesses liefern kann.

Der KF-Ansatz besteht somit aus zwei Modellgleichungen. Im Falle eines diskreten Treatments z.B. wird in einem ersten Schritt anhand eines Probit-Modells und unter Verwendung der Instrumentenvariablen die Selektion in den Treatment-Status modelliert. Aus dieser Schätzung wird ein Korrekturterm generiert, der die aus der Probitschätzung vorhersagte Korrelation des Störterms der ursprünglichen Regressionsgleichung mit dem Treatment-Status misst. Die Aufnahme des Selektionskorrekturterms in die zu schätzende Regressionsgleichung führt dazu, dass der Treatment-Status nicht mehr mit unbeobachteten Einflüssen dieser Gleichung korreliert ist und als exogen angesehen werden kann. Der KF-Ansatz ist somit eng mit dem IV-Ansatz verwandt. Allerdings gilt für den KF-Ansatz die restriktive Annahme, dass der Fehlerterm der Probit-Regression normalverteilt sein muss. Vella und Verbeek (1999) zei-

gen, dass beide Ansätze unter bestimmten Annahmen zum gleichen Ergebnis kommen.

Unter der Voraussetzung, dass die Instrumentenvariable die oben genannten Annahmen erfüllt, können sowohl der IV- als auch der KF-Ansatz eine kausale Abschätzung des Treatmenteffektes gewährleisten. Ein grundlegendes Problem bei der Anwendung eines IV- oder KF-Ansatzes im Kontext der Mindestlohnforschung besteht jedoch darin, ein gültiges Instrument zu finden. In den 1990er Jahren hat es einige Studien gegeben, die IV-Ansätze genutzt haben, um den Effekt eines Mindestlohns auf die Beschäftigung zu untersuchen, siehe z.B. Dickens et al. (1999), Machin und Manning (1994) und Neumark und Wascher (1992). Dafür nutzen die verschiedenen Autoren als Instrument zurückliegende Mindestlohnlevels bzw. zurückliegende durchschnittliche Löhne, von denen angenommen wird, dass sie stark mit dem derzeitigen Mindestlohnlevel korreliert sind, aber nicht mit dem aktuellen Beschäftigungsniveau. Letztere Annahme trifft dabei nur zu, wenn die Fehlerterme über die Zeit nicht miteinander korrelieren, wovon aber nicht ausgegangen werden kann, zumal Löhne und Mindestlöhne äußerst persistent über die Zeit sind (Blundell und Bond 1998). Daher können solche Instrumente einen Mindestlohneffekt auf die Beschäftigung nicht sauber identifizieren.

Sen et al. (2011) nutzen alternative Instrumentenvariablen, um für Kanada den Einfluss des Mindestlohns auf die Jugendarbeitslosigkeit und die Armut zu untersuchen. Die Autoren wenden den IV-Ansatz an, da sie vermuten, dass Mindestloohnerhöhungen eine politische Antwort auf eine Zunahme der Jugendarbeitslosigkeit sein könnten und daher gewöhnliche OLS-Regressionen zu Verzerrungen führen sollten. Unter der Annahme, dass linksgerichtete Parteien Mindestloohnerhöhungen eher beschließen als rechtsgerichtete Parteien verwenden die Autoren die auf der jeweiligen regionalen Ebene (Provinz) regierende politische Partei als Instrumentenvariable.¹⁷ Die zugrundeliegenden Annahmen des IV-Ansatzes lauten daher, dass (i) die regierende Partei Einfluss

¹⁷ Darüber hinaus nutzen sie als Instrumentenvariablen den regionalen Gewerkschaftsgrad sowie den Anteil an Personen, die im Handelssektor arbeiten. Beide Instrumente besitzen jedoch keine große Erklärungskraft.

auf die Entwicklung der Mindestlohnhöhe hat, während (ii) die Jugendarbeitslosigkeit außer über den Kanal der Mindestlohnhöhe nicht durch die regierende Partei beeinflusst wird. Letztere Annahme darf zumindest kritisch gesehen werden, da angenommen werden kann, dass sich die Parteien nicht nur in Bezug auf die Mindestlohnpolitik voneinander unterscheiden, sondern womöglich auch in Bezug auf andere Politikmaßnahmen, die sich wiederum unterschiedlich auf die Entwicklung der Jugendarbeitslosigkeit auswirken könnten. Für diese Politikmaßnahmen müsste gegebenenfalls kontrolliert werden.

Die Ergebnisse der IV-Schätzung weichen insgesamt kaum von denen der OLS-Regressionen ab. Damit reißen sie sich in eine Reihe von Studien ein, die finden, dass eine mögliche Verzerrung der Schätzergebnisse nur zu einem geringen Grad mithilfe von Instrumentenvariablen korrigiert werden kann (Card und Krueger 1995, Kim und Taylor 1995). Zu diesem Schluss kommt auch Lemos (2005), die ebenfalls mithilfe von politischen Instrumentenvariablen den Einfluss des Mindestlohns auf die Beschäftigung in Brasilien untersucht. Ihre Analyse ergibt, dass ihre verwendeten Politik-Instrumentenvariablen eine höhere exogene Variation aufweisen und die testbaren IV-Annahmen eher erfüllt sind, als wenn zeitlich verzögerte Mindestlohniveaus als Instrumente genutzt werden.

Dennoch bleibt für alle bisherigen Mindestlohnstudien, die den IV-Ansatz verwenden, festzuhalten, dass die zugrundeliegenden Instrumentenvariablen in Bezug auf die Einhaltung der erforderlichen IV-Annahmen nicht frei von Kritik sind. Dies verdeutlicht, wie schwierig sich in der Mindestlohnforschung geeignete Instrumentenvariablen finden lassen. Neumark und Wascher (2006) räumen ein, dass das Problem der Endogenität gerade in der Mindestlohnforschung ein wichtiges Problem ist, für das es bislang keine zufriedenstellende Lösung gibt, da die Erfüllung der IV-Annahmen für gewöhnlich nicht gegeben ist.

Vermutlich kam auch aus diesem Grund der KF-Ansatz bislang in der Mindestlohnforschung kaum zum Einsatz. Eine Ausnahme bildet die Studie von Green und Harrison (2009), die anhand kanadischer Daten untersuchen, inwiefern die Mindestlohnhöhe die Fairnessvorstellungen einer Gesellschaft reflektiert. Unter Ausnutzung der regionalen Variation von Mindestlöhnen in Kanada untersuchen die Autoren hierzu u.a. den Einfluss des durchschnittlichen Mindest-

lohnes in anderen Regionen (Provinzen) auf das jeweilige regionale Mindestlohniveau. Das durchschnittliche Mindestlohniveau anderer Provinzen ist den Autoren zufolge als endogen zu betrachten, da sich eine Provinz bei der Höhe ihres Mindestlohns an den anderen Provinzen orientieren und damit die Mindestlohnhöhe in der eigenen Provinz durch die Mindestlohnhöhe in den anderen Provinzen beeinflusst sein dürfte. Daher verwenden sie als Instrumentenvariablen die ideologischen Positionen der jeweiligen die Mindestlohnhöhe bestimmenden Regierungen der Restprovinzen. Die Begründung für die Wahl dieser Instrumente lautet, dass diese Variablen mit der Mindestlohnhöhe in den Restprovinzen stark korrelieren, während sie die Mindestlohnhöhe in der jeweilig betrachteten Provinz nicht bzw. nur über den Kanal der Mindestlohnhöhe der Restprovinzen beeinflussen sollten. Ähnlich wie beim IV-Ansatz kann diese Annahme nicht getestet werden und ist wie bei den bereits genannten Instrumenten kritisch zu bewerten.

Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne: Da die Mindestlohnregelungen nach dem AEntG bislang durch den Bund und nicht durch die Länder erstreckt werden, kommt eine Identifikationsstrategie durch Nutzung von politischen Parteien als Instrumentenvariable für Deutschland nicht in Frage. Die alternative Verwendung zeitlich verzögerter Mindestlohnlevels erscheint gerade vor dem Hintergrund der diskutierten Politikendogenität ebenfalls als höchst problematisch. Insofern ist fraglich, ob im deutschen Kontext ein geeignetes Instrument gefunden werden kann, welches hilft, Mindestlohneffekte kausal zu identifizieren.

4 Methodische Ansätze zur Ex-Ante-Evaluation (Strukturelle Modelle)

Die bisherigen Ansätze zur Ex-Post-Evaluation untersuchen die Folgen einer tatsächlich beobachteten Einführung oder Veränderung von Mindestlöhnen. Sie ermitteln die Mindestlohnwirkungen durch den Vergleich von Treatment- und Kontrollgruppen – letztere sind annahmegemäß von der Variation im Mindestlohn weder direkt noch indirekt betroffen. Die Ex-Post-Ansätze verzichten auf eine Modellierung der generellen Wirkungszusammenhänge und fokussieren bewusst auf die Evaluation einer bestimmten Politikvariation. Im Mittelpunkt der Analyse stehen die zentralen Ergebnisvariablen der Politik, etwa die Beschäftigung in einer Branche. Die Wirkungskanäle werden vor allem dann thematisiert, wenn sie für eine saubere Identifikation des kausalen Gesamteffekts von Bedeutung sind.

Die strukturellen Modelle, die in den folgenden Kapiteln vorgestellt werden, setzen es sich demgegenüber zur Aufgabe, die zentralen Wirkungszusammenhänge explizit zu modellieren. Sie treffen also Annahmen über die Ziele der relevanten Akteure (hier vor allem der Arbeitnehmer und Unternehmen) und die Restriktionen, unter denen sie handeln. Im vorliegenden Zusammenhang wird also beispielsweise modelliert, wie die beiden Marktseiten zusammenwirken, um Stellen zu schaffen und zu besetzen, und welche Informationen und Erwartungen die Akteure hinsichtlich dieses Prozesses haben. Aus den Zielen und Restriktionen werden die optimalen Handlungen der Akteure hergeleitet. Ändern sich die Rahmenbedingungen unter dem Einfluss der Politik oder aus anderen Gründen, dann werden sich die Marktteilnehmer in der Regel anders verhalten. Es ergibt sich dann ein neues Gleichgewicht, das etwa mit einem anderen Beschäftigungsniveau verbunden ist.

Der strukturelle Ansatz hat prinzipiell drei Vorteile: Zum einen bieten die Modelle einen Rahmen zur Interpretation der Gesamtwirkung. Sie liefern nämlich nicht nur Aussagen zur Wirkung eines Mindestlohns auf die gesamte Beschäftigungshöhe, sondern auch über die zugrundeliegenden Wirkungsmechanismen und die Merkmale der betreffenden Märkte. Letztere sind häufig von intrinsischem Interesse: So können durch die Schätzung struktureller Modelle beispielsweise Aussagen über die Marktmacht der Unternehmen hinsichtlich

der Arbeitsnachfrage (mit den Grenzfällen Monopson und vollkommener Wettbewerb) getroffen werden, die in der Diskussion um die Mindestlohnwirkungen eine wichtige Rolle spielt (vgl. Möller 2012, S. 198).

Zum anderen kann die einmal modellierte Struktur dazu verwendet werden, die Wirkungen bisher noch nicht beobachteter Rahmenbedingungen zu simulieren. Das ist besonders für die Ex-Ante-Abschätzung von Politikreformen von Interesse. Es kann also zum Beispiel im Vorhinein simuliert werden, welche Folgen die Einführung eines Mindestlohns oder die Erhöhung eines bestehenden Branchenmindestlohns hätte. Es lässt sich auch abschätzen, ob ein Mindestlohn eine andere Wirkung entfalten würde, wenn es gelänge, beispielsweise die Marktstruktur einer Branche zu verändern. Weiterhin kann ein struktureller Ansatz auch dazu verwendet werden, Aussagen über Mindestlohnwirkungen in bisher noch nicht betroffenen Branchen zu treffen. Zentrale Voraussetzung hierfür ist die korrekte Modellierung der dortigen Wirkmechanismen und Branchenstrukturen. Unter der Annahme, dass der Mindestlohn zwar einen Eingriff in die Lohnfindung darstellt, die übrigen Mechanismen und Parameter (also z.B. die Produktionstechnologie der Firmen, die Präferenzen der Arbeitssuchenden für Freizeit und Konsum) aber unverändert lässt, kann die so ermittelte Struktur dazu benutzt werden, um die Wirkungen eines Mindestlohnes in einer noch nicht betroffenen Branche vorherzusagen.

Drittens schließlich erlauben die Annahmen über die Zielfunktionen der Akteure eine Wohlfahrtsbewertung von Politikreformen. Die Parameter der Zielfunktionen sind in einem ersten Schritt aufgrund des beobachteten Verhaltens zu schätzen und liefern Informationen darüber, welches Gewicht die Arbeitnehmer ihrer Freizeit einerseits und zusätzlichem Konsum andererseits beimessen. Damit kann untersucht werden, in welchem Maße eine Politikreform, die beispielsweise die Beschäftigung auf Kosten der Löhne erhöht, die Arbeitnehmer tatsächlich besserstellt. Eine reine Analyse mit den in Kapitel 3 vorgestellten Ex-Post-Methoden trifft hingegen lediglich Aussagen über die Wirkungen auf Beschäftigung und Löhne, erlaubt aber keine Aussagen darüber, wie die Marktteilnehmer diese Wirkungen bewerten.

Diesen Vorteilen struktureller Modelle stehen jedoch auch Nachteile gegenüber: Die zusätzlichen Aussagen über die Wirkungszusammenhänge und das Verhalten in kontrafaktischen Situationen lassen sich nur durch zusätzliche

Annahmen gewinnen.¹⁸ Die auf diesen Annahmen aufbauende Herleitung der optimalen Reaktionen und des Gleichgewichtszustands ist zudem formal aufwändig und für Außenstehende oft schwierig nachzuvollziehen. Hinzu kommt eine in der Regel rechenintensive Anpassung des Modells an die Daten. Ex-Post-Methoden sind hier in der Schätzung deutlich einfacher und transparenter: Im einfachsten Fall können die Parameter durch die Berechnung zweier Differenzen und ihren anschließenden Vergleich ermittelt werden. Zudem erlauben es Ex-Post-Methoden prinzipiell in einfacher Form die Heterogenität der Effekte der Mindestlohnregelungen für verschiedene Arbeitnehmergruppen und Unternehmen zu schätzen.

Die Unterschiede zwischen den Ex-Post-Schätzungen und den strukturellen Modellen sollten aber nicht im Sinne eines scharfen Gegensatzes verstanden werden. Die Ansätze ergänzen sich vielmehr: Die Vorhersagen der strukturellen Modelle sollten, wo immer möglich, anhand von quasi-experimentellen Studien validiert werden, um die Glaubwürdigkeit der komplizierten Modelle zu erhöhen. Umgekehrt können die strukturellen Modelle die Basis für eine Interpretation und Verallgemeinerung der Ergebnisse von Ex-post-Studien bilden.

Auf die Bedeutung der Wirkungskanäle für die Interpretation der Ergebnisse von Ex-Post-Studien ist in den bisherigen Kapiteln schon mehrfach im Zusammenhang mit einer möglichen Verzerrung der geschätzten Effekte aufgrund von Skalen- und Substitutionseffekten in der Arbeitsnachfrage verwiesen worden. Diese Effekte ließen sich prinzipiell innerhalb eines strukturellen Arbeitsnachfragemodells quantifizieren, welches zudem berücksichtigt, dass die bislang weitgehend separat betrachteten Zielgrößen sich simultan und nicht unabhängig voneinander verändern dürften. Zur quantitativen Abschätzung der

¹⁸ Allerdings ist zur Verteidigung der strukturellen Modelle anzumerken, dass diese Annahmen der Preis für weitergehende Aussagen sind. Auch bei der Interpretation von Ex-Post-Ergebnissen besteht eine Tendenz zur Verallgemeinerung der Ergebnisse über den Kontext hinaus, in dem sie gewonnen wurden. Diese Verallgemeinerung erfolgt dann aber implizit auf Basis von „Erfahrung“ oder „Intuition“. Die strukturellen Modelle machen die Grundlagen, auf denen die Ergebnisse verallgemeinert werden, hingegen explizit und damit einer Überprüfung zugänglich.

Bedeutung von Substitutionseffekten könnten Informationen über Verschiebungen der Lohnstruktur und der relativen Beschäftigungsentwicklung aus der Vergangenheit genutzt werden, während die Bedeutung von Skaleneffekten aus Beschäftigungseffekten in Reaktion auf allgemeine Lohnkostenanstiege ohne Verschiebung der Lohnstruktur abgeschätzt werden könnte (siehe z.B. Hamermesh 1993). Die Skaleneffekte hängen insbesondere von der Gütermarktstruktur (Elastizität der Güternachfrage bei monopolistischer Konkurrenz) ab. Über den Skaleneffekt besteht somit ein direkter Zusammenhang zwischen der Wettbewerbsstruktur auf den Absatzmärkten und den Arbeitsnachfrageeffekten.¹⁹ Unserer Kenntnis nach gibt es bislang keine empirische Ex-Post-Studie, die die Verzerrung der geschätzten Effekte aufgrund von Skalen- und Substitutionseffekten mit Hilfe eines strukturellen Arbeitsnachfragemodells vorgenommen hat. Eine solche Analyse müsste unterstellen, dass auf dem Arbeitsmarkt keine monopolistische Konkurrenz vorliegt.

Eine weitere wichtige Modellklasse struktureller ökonometrischer Modelle sind Suchmodelle. Diese sind für die Evaluation von Mindestlohnwirkungen von besonderer Relevanz, da sich hieraus Rückschlüsse über die Marktmacht von Unternehmen am Arbeitsmarkt ziehen lassen. Aufgrund ihrer erheblichen Komplexität werden diese in den folgenden Abschnitten ausführlicher vorgestellt.

4.1 Strukturelle ökonometrische Schätzung von Suchmodellen

Zur Darstellung struktureller Suchmodelle wird im Folgenden zunächst kurz die Modellierung des Suchverhaltens der Arbeitnehmer im partiellen Gleichgewicht dargestellt. Diese Modelle sind Ausgangspunkt und wichtiger Baustein der Betrachtung im allgemeinen Gleichgewicht, greifen aber an einigen Stellen zu kurz und werden daher um die Modellierung der Unternehmensseite er-

¹⁹ Ein Beispiel: Wenn die Güternachfrage auf Branchenebene nahezu unelastisch ist, dann wird es im Ergebnis trotz eines gegebenenfalls intensiven Wettbewerbs zwischen den Unternehmen in einer Branchen möglich sein, den durch die Mindestlohnregelungen ausgelösten Lohnkostenanstieg eins-zu-eins auf die Absatzpreise zu überwälzen. In einem solchen Fall wäre der Skaleneffekt vernachlässigbar.

gänzt. Darauf folgen die Diskussion der Wirkungen eines Mindestlohns in Suchmodellen, die beide Marktseiten abbildet, und die Skizzierung der Anpassung der Modelle an die Daten. Anschließend folgt eine Zusammenfassung der Ergebnisse empirischer Studien zu den Mindestlohnwirkungen. Das Kapitel schließt mit Überlegungen zu einem auf den deutschen Kontext zugeschnittenen Suchmodell im allgemeinen Gleichgewicht. Weiterführende Darstellungen finden sich in den Übersichtsartikeln von van den Berg (1999), Eckstein und van den Berg (2007), Chowdry et al. (2009) und Garloff (2010). Die beiden letztgenannten Arbeiten gehen auch speziell auf die Rolle von Suchmodellen bei der Evaluation von Mindestlohnwirkungen ein.

4.1.1 Hintergrund: Suchmodelle im partiellen Gleichgewicht

Die ersten Suchmodelle auf dem Arbeitsmarkt beschränkten sich auf das Verhalten der Arbeitsanbieter.²⁰ Der Fokus lag dabei auf den derzeit nicht Beschäftigten. Diese erhalten von Zeit zu Zeit Stellenangebote und entscheiden, ob sie diese annehmen. Die Arbeitsanbieter wissen lediglich, wie häufig die Angebote eingehen und mit welcher Wahrscheinlichkeit sie Angebote einer bestimmten (Lohn-)Höhe erhalten, treffen ihre Entscheidung also unter Unsicherheit. Stellenangebote mit einem Lohn der so niedrig ist, dass er sie nicht für den Wegfall an frei verfügbarer Zeit entschädigt, werden sie auf jeden Fall ablehnen. Bei höheren Löhnen unterliegen sie einem Zielkonflikt: Einerseits können sie sich dadurch, dass sie die Stelle annehmen, sofort verbessern. Andererseits bedeutet die Annahme den Verzicht auf zukünftige Angebote mit möglicherweise höherem Lohn.

Unter diesen Annahmen lässt sich ein sogenannter Anspruchslohn herleiten, mit dem sich das (optimale) Verhalten charakterisieren lässt. Der Anspruchslohn ist ein Schwellenwert: Stellenangebote mit niedrigerem Lohn werden abgelehnt, Angebote mit höherem Lohn werden angenommen. Diese Entscheidungsregel ist keine Garantie dafür, dass sich die Arbeitsanbieter in je-

²⁰ Im Sprachgebrauch der Ökonomie sind es die *Arbeitnehmer*, die Arbeit anbieten. Die Firmen fragen die Arbeitsleistungen nach, bilden also die Nachfrageseite. Modelle im partiellen Gleichgewicht beschreiben nur eine der beiden Marktseiten; im allgemeinen Gleichgewicht wird ihr Zusammenspiel modelliert.

dem einzelnen Fall im Rückblick am besten stellen – das hängt ja von den konkreten Stellenangeboten ab, die niemand im Voraus kennen kann. Das am Anspruchslohn orientierte Verhalten stellt aber zumindest sicher, dass sich die Arbeitsanbieter unter den ihnen zum Entscheidungszeitpunkt zur Verfügung stehenden Informationen optimal verhalten.

Die genaue Höhe des Anspruchslohns hängt unter anderem davon ab, wie häufig die Stellenangebote eingehen und wie „ungeduldig“ die Arbeitsanbieter sind, wie stark sie also gegenwärtige Periode im Vergleich mit zukünftigen Perioden gewichten. Daneben spielt die statistische Verteilung der Lohnangebote eine wichtige Rolle. Für die Wirtschaftspolitik schließlich ist interessant, welchen Einfluss die Dauer und Höhe der Arbeitslosenunterstützung auf das Suchverhalten und damit die Dauer der Arbeitslosigkeit ausüben. Hierzu wurden seit den 1970er Jahren zahlreiche empirische Studien vorgelegt – Tatsiramos und van Ours (2012) und van den Berg und Uhlendorff (2013) liefern einen aktuellen Überblick über die umfangreiche Literatur.

4.1.2 Grenzen des Partialansatzes

Die Arbeitsnachfrageseite wird im partiellen Suchmodell durch die Häufigkeit der Stellenangebote und die Verteilung der Lohnhöhe beschrieben. Beide werden als gegeben angenommen, reagieren im partiellen Modell also nicht auf eine Politikveränderung wie beispielsweise die Einführung oder Variation eines Mindestlohns oder eine Anpassung der Arbeitslosenunterstützung. Das ist für die Politikanalyse ungünstig, weil bestimmte Wirkungskanäle und Anpassungsmechanismen dadurch von vornherein ausgeschlossen werden. Auf diesen Punkt wird unten noch einmal ausführlich eingegangen.

Hinzu kommt ein theoretischer Kritikpunkt: Im partiellen Suchmodell wird eine echte *Verteilung* der Lohnangebote angenommen – hätten alle Angebote die gleiche Höhe, dann bräuchten die Arbeitnehmer nicht zu suchen. Wie Diamond (1971) gezeigt hat, ist die Annahme aber problematisch, da in Suchmodellen die Unternehmen keinen Anreiz haben, etwas anderes als den Anspruchslohn anzubieten. Bieten sie weniger, wird niemand für sie arbeiten; bieten sie mehr, dann steigen die Kosten, ohne dass die Unternehmen zusätzliche Arbeitnehmer für sich gewinnen können. Daraus folgt: Gibt es nur *einen einzigen* Anspruchslohn, dann haben alle Lohnangebote die gleiche Höhe. Die

„Verteilung“ besteht also nur aus einem einzigen Punkt. Damit wird aber die Suche der Arbeitnehmer hinfällig.

Die Modelle im allgemeinen Gleichgewicht stellen auch eine Reaktion auf diese Kritik dar und müssen sich daran messen lassen, ob sie aus dem Verhalten der Arbeitgeber eine echte Lohnverteilung herleiten können. Albrecht und Axell (1984) zeigen, dass diese Herleitung gelingt, wenn sich die derzeit Nichtbeschäftigten in ihrem Anspruchslohn unterscheiden, etwa weil sie unterschiedlich viel Arbeitslosenunterstützung erhalten. Die Annahme ist plausibel und führt zu einer echten Lohnverteilung. Eckstein und Wolpin (1990) zeigen allerdings, dass eine diskrete Anzahl von Anspruchslöhnen der Nichtbeschäftigten nicht ausreicht, um die tatsächlich beobachtete stetige Lohnverteilung zu reproduzieren bzw. dass die Reproduktion nur dann gelingt, wenn man unterstellt, dass die beobachteten Löhne mit erheblichen Fehlern gemessen wurden.²¹

Eine deutliche Verbesserung wird im Modell von Burdett und Mortensen (1998) dadurch erreicht, dass auch bereits beschäftigte Arbeitnehmer weiterhin nach neuen Jobs Ausschau halten („on-the-job search“). Ihr Anspruchslohn entspricht ihrem derzeitigen Lohn. Im Gleichgewicht ergibt sich selbst dann eine stetige Lohnverteilung, wenn die Nichtbeschäftigten alle den gleichen Anspruchslohn haben. Die Variation der Anspruchslöhne wird also im Modell generiert. Das ist eleganter als der Ansatz von Albrecht und Axell (1984), die die Existenz verschiedener Anspruchslöhne von vornherein unterstellen. Außerdem lässt sich mit dem Burdett-Mortensen-Modell die tatsächliche, stetige Lohnverteilung deutlich besser reproduzieren, wie von van den Berg und Ridder (1998) für die Niederlande gezeigt wurde.²²

²¹ Wie oben ausgeführt, werden die Anspruchslöhne im Modell selbst hergeleitet. Wenn also hier und im Folgenden von einer diskreten Zahl (oder generell: einer Verteilung) von Anspruchslöhnen die Rede ist, so ist dies lediglich eine Kurzform dafür, dass das Modell für verschiedene Typen von Arbeitnehmern unterschiedliche Anspruchslöhne *generiert*.

²² Viele Suchmodelle im allgemeinen Gleichgewicht generieren allerdings eine rechtssteile Lohnverteilung. Die in der Realität beobachtete linkssteile, rechtsschiefe Lohnverteilung erhält man in diesen Modellen typischerweise nur dadurch, dass man eine entsprechende Verteilung der Anspruchslöhne und/oder der Produktivität unterstellt. Eine Alternative

4.1.3 Modellierung des Arbeitgeberverhaltens

Die Arbeitsnachfrageseite besteht in Suchmodellen im allgemeinen Gleichgewicht aus einer Vielzahl von Firmen, die in jeder Periode Stellen zu einem von ihnen festgelegten Lohnsatz ausschreiben. Dieser Lohnsatz ist nicht verhandelbar („wage posting“).²³ Die Unternehmen sehen sich einem konstanten Güterpreis gegenüber und verwenden im einfachsten Modellfall alle die gleiche lineare Produktionstechnologie mit Arbeit als dem einzigen Inputfaktor.

Das aus Unternehmenssicht optimale Lohnangebot ergibt sich aus einem Gewinnmaximierungskalkül. Wie oben dargelegt, ist ein Lohn unterhalb des niedrigsten Anspruchslohns ebenso wenig sinnvoll wie ein Lohn, der über dem höchsten Anspruchslohn liegt. Dazwischen besteht ein Zielkonflikt: Je höher das Lohnangebot, desto niedriger ist einerseits der Gewinn pro Arbeitseinheit. Andererseits kann ein Unternehmen durch einen höheren Lohn aber die Belegschaft und damit den Output vergrößern, weil mehr Arbeitsuchende die Lohnangebote akzeptieren werden. Außerdem können bei einem höheren Lohn Beschäftigte der Konkurrenz abgeworben werden, und das Unternehmen verliert weniger eigene Mitarbeiter. Im Gleichgewicht führen die Hochlohn- und Niedriglohnstrategien zum gleichen Gewinn.

Ein Merkmal der wage-posting-Modelle im allgemeinen Gleichgewicht ist, dass Arbeitgeber über Marktmacht verfügen, also einen Lohn zahlen, der unterhalb der Grenzproduktivität liegt. Ursache der Marktmacht sind die Friktionen auf dem Arbeitsmarkt: Die Arbeitnehmer können nicht sofort den Job wechseln, sondern erst dann, wenn sie ein besseres Stellenangebot erhalten. Die Arbeitnehmer verbessern sich also im Laufe ihres Erwerbslebens tendenziell solange, bis sie bei einem Arbeitgeber arbeiten, bei dem ihr Lohn (nahezu) der Grenz-

hierzu besteht darin, die linkssteile, rechtsschiefe Lohnverteilung als Mischung rechtssteiler Lohnverteilungen für einzelne Arbeitnehmergruppen zu erzeugen. Van den Berg und Ridder (1998) wählen diesen Weg und definieren segmentierte Arbeitsmärkte. Dieser Ansatz wird später noch einmal aufgegriffen.

²³ Ein alternativer Ansatz sind die sogenannten Bargaining-Modelle, in denen die Firmen lediglich Stellen ausschreiben und Arbeitnehmer und Firmen anschließend über die Lohnhöhe verhandeln. Abschnitt 4.1.6 geht kurz auf diesen Ansatz ein.

produktivität entspricht. Im Ansatz von Burdett und Mortensen (1998) wird dieser Grenzfall aber nicht von allen Arbeitnehmern erreicht, da mit exogener Rate Arbeitsverhältnisse aufgelöst werden und die Betroffenen ihren nächsten Job dann aus der Arbeitslosigkeit und mit entsprechend niedrigerem Anspruchslohn suchen.²⁴

4.1.4 Anpassung an die Daten

Das Verhalten von Unternehmen und Arbeitnehmern ist abhängig von den Parametern des Modells. Um die Wirkungen eines Mindestlohns zu quantifizieren, muss das Modell daher zunächst an die Daten angepasst werden. Dazu werden Parameterwerte gewählt, für die das Modell den tatsächlichen Beobachtungen möglichst nahe kommt. Anschließend kann das so parametrisierte Modell auch zur Vorhersage in solchen Situationen verwendet werden, für die noch keine Beobachtungen vorliegen.

Die Schätzung des Modells stellt niedrigere Anforderungen an die Daten, als auf den ersten Blick zu vermuten wäre. Erforderlich sind Daten zur Lohnverteilung und zu den Übergängen zwischen Arbeitslosigkeit und Beschäftigung sowie, in Modellen mit „on-the-job search“, zu den Übergängen zwischen verschiedenen Jobs. Obwohl beide Marktseiten modelliert werden, sind Unternehmens-/Betriebsdaten nicht zwingend erforderlich. Diese bieten aber Vorteile bei der Schätzung komplizierter Modelle (vgl. van den Berg 1999, S. 302 und Chowdry et al. 2009, S. 15 ff.).

Zur Schätzung wird in der Regel das Maximum-Likelihood-Verfahren verwendet. Dazu werden für eine bestimmte Parameterkonstellation die Vorhersagen des Modells hinsichtlich der Lohnverteilung und der Beschäftigungs- und Arbeitslosigkeitsdauer ermittelt. Diese Vorhersagen werden dann mit den tatsächlichen Beobachtungen verglichen. Liegen beide weit auseinander, werden die Vorhersagen für neue Parameterwerte wiederholt und wieder mit den

²⁴ Die Annahme, dass die Arbeitsverhältnisse mit exogener Rate beendet werden, ist ein Defizit bei der Modellierung des Arbeitgeberverhaltens. Der Fall, dass Entlassungen durch Lohnerhöhungen induziert werden, wird durch die Annahme nämlich nicht erfasst.

Daten verglichen – solange, bis beide nah genug beieinander liegen, das Modell also die Beobachtungen mit hinreichender Genauigkeit abbildet.

Dabei ist sicherzustellen, dass die optimalen Parameterwerte eindeutig sind, also nicht mehrere Parameterkombinationen die gleichen Vorhersagen und damit die gleiche Güte der Modellanpassung liefern. Im einfachsten Fall – also bei Homogenität von Arbeitnehmern und Unternehmen – wird die *job arrival rate* aus den beobachteten Übergängen von der Arbeitslosigkeit in die Beschäftigung ermittelt, die *job destruction rate* durch Übergänge in die andere Richtung. Die Übergänge zwischen zwei Beschäftigungsverhältnissen liefern Informationen über die *job transition rate*. Die Produktivität wird aus der Lohnverteilung erschlossen. Der Nutzenwert bei Nichtbeschäftigung kann auf verschiedene Arten festgelegt werden: Zum einen lassen sich Informationen über die Höhe staatlicher Sozialleistungen und die mit ihnen verbundenen Anspruchsbedingungen verwenden (van den Berg und Ridder 1998, van Vuuren et al. 2000); alternativ kann der Anspruchslohn als der niedrigste in den Daten beobachtete Lohn definiert (Bowlus et al. 1995, 2001; Bunzel et al. 2001) oder durch direkte Fragen in sozialwissenschaftlichen Erhebungen („Wie viel müssten Sie mindestens verdienen, um eine Beschäftigung aufzunehmen?“) ermittelt werden.

Bei Heterogenität im Modell müssen zusätzliche Parameter geschätzt werden, die etwas über Art und Anteil der verschiedenen Typen im Modell aussagen. Das genaue Verfahren ist abhängig von der Art der Heterogenität (Produktivität von Unternehmen oder Arbeitnehmern, Anspruchslöhne der Arbeitnehmer). Einzelheiten hierzu finden sich bei Chowdry et al. (2009, S. 11-12).

4.1.5 Theoretische Wirkungen eines Mindestlohns

Das Suchmodell im allgemeinen Gleichgewicht enthält zwei für die Diskussion der Mindestlohnwirkungen wichtige Modelle als Grenzfälle (vgl. Cahuc und Zylberberg 2004): Je häufiger die Stellenangebote eingehen, desto geringer die Marktmacht der Unternehmen und desto näher ist das Suchmodell am Modell des vollkommenen Wettbewerbs, in dem jeder Arbeitnehmer nach seiner Grenzproduktivität entlohnt wird. Sind die Friktionen hingegen hoch, dann steigt die Marktmacht der Arbeitsnachfrager. Das Monopsonmodell kann daher ebenfalls als Grenzfall des Suchmodells im allgemeinen Gleichgewicht betrachtet werden. Die Suchfriktionen liefern die mikroökonomische Unterfütte-

rung der Marktmacht, die ansonsten im statischen Lehrbuchmodell des Monopsons lediglich angenommen wird. Die Stärke der Suchfraktionen und damit die Marktmacht können mithilfe des Modells empirisch ermittelt werden. Dies ist ein großer Vorteil des Ansatzes, da die Größe der Marktmacht auch für die Interpretation der Ergebnisse quasi-experimenteller Studien zum Mindestlohn von Bedeutung ist.

Die Wirkungen eines Mindestlohns in einem Suchmodell im allgemeinen Gleichgewicht sind abhängig von seiner Höhe und Ausgestaltung, von der Marktmacht der Unternehmen und davon, welche Art der Heterogenität berücksichtigt wird.

Im einfachsten Fall eines allgemeinen Mindestlohns in einem Modell **ohne Heterogenität** (also mit identischer Produktivität der Unternehmen und identischer Produktivität und Opportunitätskosten der Arbeitnehmer) sind drei Konstellationen zu unterscheiden. Liegt der Mindestlohn unterhalb des niedrigsten derzeitigen Marktlohns, dann ist er unwirksam. Liegt er höher, dann ist die Beschäftigungswirkung abhängig davon, ob der Gleichgewichtslohn der Grenzproduktivität entspricht. Haben die Arbeitgeber aufgrund der Suchfraktionen Marktmacht und entlohnen sie die Arbeitnehmer unterhalb ihrer Grenzproduktivität, dann führt ein Mindestlohn zunächst zu einer Umverteilung zugunsten der Arbeitnehmer, ist aber beschäftigungsneutral. Ein Mindestlohn, der über der Grenzproduktivität liegt, geht allerdings auf Kosten der Beschäftigung – im hier betrachteten homogenen Fall scheiden dann sogar *alle* Unternehmen aus dem Markt aus. Im einfachsten, homogenen Fall kostet ein Mindestlohn also entweder alle Arbeitsplätze oder gar keinen.

Heterogenität hinsichtlich der Produktivität kann auf zwei Arten modelliert werden. Ist die Produktivität eine Eigenschaft der Unternehmen, dann sinkt die Beschäftigung, sobald der Mindestlohn über das Produktivitätsniveau des Unternehmens mit der geringsten Produktivität hinaus erhöht wird. Alternativ kann die Produktivität als Eigenschaft der Arbeitnehmer modelliert werden (van den Berg und Ridder 1998). Dazu wird eine Vielzahl von annahmegemäß segmentierten Arbeitsmärkten betrachtet, die zum Beispiel anhand von Alter oder Bildung definiert werden; innerhalb eines Marktes wird die Produktivität dann wieder als homogen betrachtet. Ein allgemeiner Mindestlohn ist dann ohne Beschäftigungswirkung, solange er niedriger liegt als die Produktivität

der am wenigsten produktiven Arbeitnehmer. In beiden Varianten sagen Modelle mit Heterogenität in der Produktivität weniger abrupte Wirkungen voraus als im oben skizzierten homogenen Fall. Ist die Produktivitätsverteilung der Unternehmen stetig oder gibt es eine große Zahl von Arbeitsmarktsegmenten, dann erfolgt der Rückgang in beiden Fällen allmählich, weil immer nur das marginale Unternehmen oder das marginale Segment unprofitabel wird.

Heterogenität im Anspruchslohn, die etwa durch unterschiedlich hohe Ansprüche auf Arbeitslosenunterstützung durch den Staat oder durch Familie und Freunde entstehen kann, eröffnet die interessante Möglichkeit, dass in einem Suchmodell durch den Mindestlohn die Beschäftigung sogar *steigen* kann. Personen mit hohen Anspruchslöhnen bleiben nämlich möglicherweise deswegen länger arbeitslos, weil sie viele Stellenangebote mit zu niedrigen Löhnen erhalten. Albrecht und Axell (1984) sprechen hier von „ineffizienter Arbeitslosigkeit“, weil es theoretisch durchaus Firmen gäbe, die den Personen mit hohen Anspruchslöhnen ein maßgeschneidertes, höheres Lohnangebot unterbreiten könnten. Ein solches selektives Eintreffen nur hoher Angebote ist aber im Modell ausgeschlossen – sämtliche Arbeitnehmer erhalten Offerten aus derselben Verteilung von Lohnangeboten. Nach Einführung eines Mindestlohns kann daher die Beschäftigung steigen, weil nun keine Lohnangebote unterhalb des Mindestlohns mehr gemacht werden dürfen, Personen mit hohem Anspruchslohn also weniger lang auf ein passendes Angebot warten müssen. Übersetzt in eine weniger modelltechnische Sprache besagt dieser Zusammenhang, dass durch einen höheren Mindestlohn nicht nur unproduktive Arbeitsplätze verloren gehen können (vgl. den Fall „Heterogenität in der Produktivität“), sondern dass zugleich die auch bei höherem Lohn noch profitablen Arbeitsplätze schneller besetzt werden können.

Modelle, die **mehrere Arten von Heterogenität** kombinieren, sind flexibel genug, um alle der skizzierten Wirkungszusammenhänge abbilden zu können. Ob ein Mindestlohn beschäftigungsneutral ist, Arbeitsplätze kostet oder im Gegenteil die Beschäftigung erhöht, ist dann eine empirische Frage. Diese Flexibilität des Ansatzes ist eine Stärke für die Evaluation, da eine Wirkungsanalyse, die das Vorzeichen von vornherein festlegt, unbefriedigend ist. Dass die Modelle in ihrer allgemeinen Form hinsichtlich des Vorzeichens und der Stärke der Beschäftigungswirkungen flexibel sind, bedeutet aber nicht, dass die Vor-

hersagen beliebig sind. Sobald aus einer Schätzung konkrete Parameterwerte vorliegen, sind die Modelle zur Vorhersage einsetzbar.

4.1.6 Bargaining-Modelle

Die bisherige Modellbeschreibung und Diskussion der Mindestlohnwirkungen hat sich auf die Klasse der „wage-posting“-Modelle konzentriert, in denen die Firmen über die Lohnhöhe entscheiden und die (potenziellen) Arbeitnehmer dann nur vor der Wahl stehen, das Stellenangebot anzunehmen oder abzulehnen. Eine zweite Klasse von Suchmodellen („**Bargaining-Modelle**“) nimmt stattdessen an, dass die Unternehmen Stellen ausschreiben und der Lohn erst anschließend in einer Verhandlung zwischen der Firma und dem Arbeitnehmer festgelegt wird. Es kommen nur solche „Matches“ zustande, die für beide Seiten vorteilhaft sind: Die Produktivität muss also mindestens dem Lohnniveau entsprechen, und der Arbeitnehmer muss mindestens seinen Anspruchslohn verdienen. Welcher Lohn innerhalb dieses Bereichs dann tatsächlich vereinbart wird, hängt von der Verhandlungsmacht beider Parteien ab, die im Modell als exogener Parameter vorgegeben ist.

Wie produktiv ein Arbeitnehmer in einem bestimmten Unternehmen ist, stellt sich erst heraus, nachdem beide Marktseiten zueinander gefunden haben. Beide Parteien müssen dann entscheiden, ob sie ein Arbeitsverhältnis eingehen oder aber weitersuchen möchten. Grundlage für diese Entscheidung ist die Kenntnis darüber, wie die Produktivität der „Matches“ statistisch verteilt ist. Diese Verteilung entscheidet auch darüber, ob es sich für die Unternehmen überhaupt lohnt, eine Stelle auszuschreiben.²⁵

In einem einfachen Bargaining-Modell hat ein nicht-bindender Mindestlohn keine und ein höherer Mindestlohn eine durchgängig negative Wirkung auf die Beschäftigung, da weniger produktive „Matches“ nun aus Arbeitgebersicht nicht mehr rentabel sind. Ahn et al. (2011) erweitern das Modell um die Entscheidung, überhaupt am Arbeitsmarkt zu partizipieren – je höher der Mindestlohn, desto mehr Personen treten auf den Arbeitsmarkt. In dieser Model-

²⁵ Auf die Schätzung von Bargaining-Modellen wird hier nicht näher eingegangen. Chowdry et al. (2009, S. 12) nennen die wichtigsten Aspekte.

erweiterung ist die Beschäftigungswirkung eines Mindestlohns in ihrem Vorzeichen a priori unbestimmt: Einerseits sinkt die Wahrscheinlichkeit, dass ein gegebener „Match“ zu einem Arbeitsverhältnis führt, andererseits steigt die Zahl der „Matches“.

4.1.7 Empirische Evidenz

Es liegt eine Reihe von empirischen Studien vor, welche die Wirkungen eines Mindestlohns im Rahmen eines Suchmodells im allgemeinen Gleichgewicht analysieren. Nur wenige Arbeiten nutzen jedoch eine Modellvariante, in der die Beschäftigungswirkungen in ihrer Richtung nicht von vornherein feststehen.

Die erste Ausnahme ist die Studie von Eckstein und Wolpin (1990) auf Grundlage amerikanischer Daten. Sie verwenden den theoretischen Rahmen von Albrecht und Axell (1984), also ein wage-posting-Modell, in dem jedoch nur das Suchverhalten derzeit nicht beschäftigter Personen abgebildet wird, in ihrem Fall von Schulabgängern – „on-the-job search“ wird nicht berücksichtigt. Eckstein und Wolpin haben daher Schwierigkeiten, die tatsächliche Lohnverteilung mit ihrem Modell zu reproduzieren. Entsprechend sehen sie ihre Simulationsergebnisse auch ausdrücklich als Illustration des grundsätzlich Möglichen und weniger als Beitrag zu einer echten Evaluation des Mindestlohns. Ihr Ergebnis, dass ein Mindestlohn zu Beschäftigungseinbußen führt, sollte daher entsprechend vorsichtig interpretiert werden.

Aussagekräftiger ist eine Studie von Bontemps et al. (1999) für Frankreich. Sie verwenden den Modellrahmen von Burdett und Mortensen (1998), der auch „on-the-job search“ abbildet und damit die Anpassungsgüte des Modells verbessert. Bontemps et al. erweitern das Modell um die Heterogenität sowohl der Produktivität als auch der Anspruchslöhne²⁶; damit ist die Richtung des Mindestlohneffekts theoretisch unbestimmt. Simulationen zeigen, dass ein niedriger Mindestlohn im Modell zunächst die Beschäftigung erhöht, weil nun

²⁶ Genauer: um Heterogenität in den Faktoren (Präferenzen, Arbeitslosenunterstützung), die dafür sorgen, dass sich Arbeitnehmer unter sonst gleichen Umständen in den im Modell generierten Anspruchslöhnen unterscheiden; siehe den Hinweis oben.

auch Personen mit hohem Anspruchslohn häufiger Stellenangebote mit für sie ausreichend hohen Löhnen erhalten. Ab einer bestimmten Höhe dominiert dann aber der negative Beschäftigungseffekt des Mindestlohns, der dadurch entsteht, dass Unternehmen mit niedriger Produktivität aus dem Markt verschwinden.

Das einzige Bargaining-Modell, das prinzipiell positive wie negative Beschäftigungswirkungen erzeugen kann, stammt von Ahn et al. (2011). Die Autoren finden, dass für Teenager in den USA der Mindestlohn insgesamt eine zu vernachlässigende Beschäftigungswirkung hat. Zwar kommen einige „Matches“ nicht mehr zustande, weil ihre Produktivität nicht hoch genug ist für den Mindestlohn; gleichzeitig treten aber mehr Jugendliche auf den Arbeitsmarkt. Beide Wirkungen sind in etwa gleich stark.

Die übrigen empirischen Studien auf Grundlage eines Suchmodells im allgemeinen Gleichgewicht unterstellen implizit, dass Mindestlöhne entweder keine oder negative Beschäftigungswirkungen entfalten. Das ist aus Sicht der Evaluation nicht ideal. Andererseits gilt, dass auch in quasi-experimentellen Studien deutlich positive Beschäftigungseffekte eher selten sind. Außerdem ist zu beachten, dass die meisten empirischen Suchmodelle nicht in erster Linie zur Evaluation von Mindestlöhnen entwickelt wurden, sondern in vielen Fällen die grundlegende Entwicklung von Modellen und Schätzverfahren im Vordergrund stand.

Die Arbeit von van den Berg und Ridder (1998) ist bemerkenswert, weil sie die erste empirische Umsetzung von Burdett und Mortensens (1998) Modell mit „on-the-job search“ darstellt. Van den Berg und Ridder lassen Heterogenität hinsichtlich der Produktivität der Arbeitnehmer zu und unterteilen den Arbeitsmarkt dazu in verschiedene Segmente, die nach Alter, Ausbildung und Art der Tätigkeit abgegrenzt sind. Die Unterscheidung dieser Typen von Arbeitnehmern ist wichtig, da in den Niederlanden der Mindestlohn altersabhängig ist. Ihre Simulationen ergeben, dass der Mindestlohn zu Arbeitslosigkeit führt, die Wirkung aber nicht linear ist: So würde ein Anstieg des damaligen Mindestlohns um 10% die Arbeitslosigkeit um 3% erhöhen, ein um 25% höherer Mindestlohn hätte mit einer um 16% höheren Arbeitslosigkeit eine mehr als fünfmal so starke Wirkung. Die negativen Beschäftigungswirkungen fallen für junge Beschäftigte am stärksten aus, weil laut den Schätzungen von van den Berg

und Ridder für diese Gruppe die Suchfraktionen auf dem Arbeitsmarkt am geringsten sind. Das bedeutet, dass die Entlohnung nahe an der Grenzproduktivität liegt und sich der Mindestlohn daher weniger in einer Umverteilung von Unternehmen zu Arbeitnehmern, sondern vor allem in einem Rückgang der nunmehr unprofitabel gewordenen Beschäftigung niederschlägt.

Van Vuuren et al. (2000) finden für Großbritannien ebenfalls negative Beschäftigungswirkungen eines Mindestlohns, obwohl das der Studie zugrunde liegende Modell grundsätzlich auch keine oder positive Beschäftigungseffekte zulässt.²⁷ Anders als van den Berg und Ridder modellieren sie keine diskrete Zahl von Produktivitätstypen, sondern eine stetige Verteilung. Außerdem berücksichtigen sie zusätzlich die Heterogenität in den Anspruchslöhnen.

Unter den Bargaining-Studien sind noch die Arbeiten von Flinn (2006) und Flinn und Mabli (2008) zu nennen; anders als bei Ahn et al. (2011) wird aber die Möglichkeit positiver Beschäftigungswirkungen eines Mindestlohns von vornherein ausgeschlossen. Beide Studien finden auf Basis amerikanischer Daten aus der zweiten Hälfte der 1990er Jahre entsprechend negative Effekte. Gleichzeitig untersuchen die Studien aber auch die Wohlfahrtswirkungen eines Mindestlohns (vgl. die Bemerkungen am Anfang von Abschnitt 4). Sie kommen zu dem Schluss, dass ein Mindestlohn bis zu einer bestimmten Höhe aus Wohlfahrtsgesichtspunkten zunächst eine Verbesserung darstellt, da die Löhne der Beschäftigten steigen. Bei höherem Mindestlohn dominieren dann die negativen Effekte, die sich aus dem Rückgang der Beschäftigung ergeben. Gorry (2013) erweitert das Modell von Flinn (2006) um den Aspekt der Berufserfahrung. In der empirischen Anwendung von Gorrays Modell für die USA wirkt ein zu hoher Mindestlohn vor allem auf junge Arbeitnehmer negativ, da er ihre Möglichkeiten einschränkt, während einer ersten Lernphase im Austausch für zunächst niedrigere Löhne Erfahrung auf dem Arbeitsmarkt zu sammeln, die sich dann produktivitätssteigernd auswirken würde und von selbst einen Weg aus dem Niedriglohnsektor weisen könnte.

²⁷ Positive Wirkungen können dann auftreten, wenn der Mindestlohn dazu führt, dass Arbeitnehmer mit hohem Anspruchslohn häufiger ein für sie akzeptables Lohnangebot erhalten.

4.1.8 Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne

Der strukturelle Ansatz, also die explizite Modellierung genereller Wirkungszusammenhänge, ist kein Ersatz für quasi-experimentelle Methoden, sondern eine sinnvolle Ergänzung: Mit diesem Ansatz lassen sich die Mechanismen untersuchen, über die der Mindestlohn seine Wirkungen entfaltet; insbesondere wird es möglich, die Marktmacht der Unternehmen abzuschätzen. Zweitens erlaubt der strukturelle Ansatz die Vorhersage von Mindestlohnwirkungen auch über die bisher beobachtete Variation hinaus. Das heißt, man kann ex-ante simulieren, welche Wirkungen eine Mindestloohnerhöhung oder die Einführung eines Mindestlohns in weiteren Branchen hätte. Drittens schließlich trifft der strukturelle Ansatz Aussagen über die Zielfunktionen von Arbeitnehmern und Firmen und erlaubt es somit, Veränderungen im Mindestlohn aus Sicht der Wirtschaftssubjekte zu bewerten. Der strukturelle Ansatz ist dort am stärksten, wo es gelingt, die Vorhersagen der oft komplizierten Modelle anhand quasi-experimenteller Studien zu validieren – dies ist allerdings, zumindest für die hier betrachtete Klasse der Suchmodelle im allgemeinen Gleichgewicht, bisher nicht systematisch erfolgt.

Aus unserer Sicht ist ein Einsatz struktureller Methoden bei der Evaluation der deutschen Branchenmindestlöhne grundsätzlich wünschenswert und machbar. Allerdings ist zu bedenken, dass die Entwicklung und Schätzung eines Suchmodells im allgemeinen Gleichgewicht mit einem erheblichen Zeitaufwand verbunden ist. Innerhalb eines Projektrahmens von beispielsweise sechs bis zwölf Monaten wäre dies nicht zu leisten.

Dies gilt umso mehr, als bisher für Deutschland lediglich zwei Schätzungen eines Suchmodells im allgemeinen Gleichgewicht vorliegt, die aber erst an die Evaluationsaufgabe angepasst werden müssten. Launov, Wolff und Klassen (2004) und Holzner und Launov (2010) entwickeln Suchmodelle à la Burdett und Mortensen (1998) und schätzen sie auf Basis des Sozio-ökonomischen Panels (SOEP). Vor allem die Arbeit von Holzner und Launov ist für die Evaluation des Mindestlohns interessant, da sie Arbeitnehmer verschiedener Qualifikation und Produktivität modellieren und dabei auch mögliche Substitutionsbeziehungen berücksichtigen. Das ist eine Erweiterung gegenüber dem oben skizzierten Ansatz segmentierter Arbeitsmärkte, zwischen

denen keine Wechselbeziehungen herrschen. Holzner und Launov zeigen, dass durch die Annahme einer gemeinsamen Produktionsfunktion das Modell eine Lohnverteilung vorhersagen kann, die der tatsächlich beobachteten, näherungsweise log-normalen Verteilung nahe kommt – bisherige Studien hatten das Problem, dass sich aus dem Modell eine Verteilung ergibt, bei der hohe Löhne häufiger auftreten als niedrige. Holzner und Launov nutzen ihr Modell für eine Untersuchung der Bildungsentscheidung; der Mindestlohn spielt in ihrem Artikel keine Rolle, ließe sich aber grundsätzlich integrieren.

Allerdings wäre dann sorgfältig zu prüfen, ob das Modell in der derzeitigen Form bereits flexibel genug ist, um sämtliche Wirkungsrichtungen des Mindestlohns auf die Beschäftigung zu erfassen. Wie oben dargelegt ergeben sich in einem homogenen Modell oder einem Modell mit Heterogenität in der Produktivität entweder keine oder negative Beschäftigungswirkungen. Um auch die Möglichkeit positiver Wirkungen zu erfassen, ist zusätzliche Heterogenität in den Anspruchslöhnen erforderlich (vgl. Bontemps et al. 1999). Modelle mit Heterogenität in beiden Dimensionen stellen höhere Datenanforderungen. In Deutschland liegt mit den verknüpften Arbeitnehmer-Arbeitgeber-Daten des IAB aber ein Datensatz vor, der hier prinzipiell geeignet erscheint.

Eine zweite offene Frage betrifft die Eignung des Modells für die Evaluation der deutschen Branchenmindestlöhne. Hier erscheint eine Kombination von Holzner und Launovs Produktionsfunktion, über die sich Substitutionsbeziehungen zwischen verschiedenen Qualifikationsgruppen erfassen lassen, mit dem Ansatz segmentierter Arbeitsmärkte von van den Berg und Ridder (1998) interessant. Über letzteren ließen sich die Branchen abbilden; Holzner und Launovs Ansatz erlaubte dann eine Untersuchung von Spillovereffekten innerhalb der Branchen.

Nicht berücksichtigt blieben allerdings mögliche Spillover-Effekte *zwischen* den Branchen. Hierfür müsste eine kompliziertere Produktionsfunktion in das Modell integriert werden. Auch die Mobilität von Arbeitnehmern zwischen den Branchen wäre abzubilden, da Arbeitnehmer und vor allem Arbeitslose Jobs in verschiedenen Branchen suchen können. Die Relevanz dieser Problematik ist vermutlich für die Gebäudereinigung, die Wäschereidienstleistungen und die Abfallwirtschaft stärker als für das Bauhauptgewerbe und zwei der drei Nebengewerbe (Dachdecker, Elektrohandwerk) Für die Umsetzung im deut-

schen Kontext der Branchenmindestlöhne kommt noch hinzu, dass letztere durch die Allgemeinverbindlicherklärung von Tarifverträgen eingeführt werden. Anders als bei einem durch den Staat festgelegten allgemeinen Mindestlohn sind diese Branchenmindestlöhne also Ergebnis eines komplexen Verhandlungsprozesses, der gegebenenfalls im Rahmen eines Bargaining-Ansatzes spieltheoretisch zu modellieren wäre.

Erweiterungen dieser Art würden den Modellierungsaufwand noch einmal deutlich steigern. Bereits die genaue Prüfung dieser Optionen nähme einige Zeit in Anspruch. An einen kurzfristigen Einsatz in der Evaluation der deutschen Branchenmindestlöhne ist daher nicht zu denken.

Angesichts des hohen Entwicklungsaufwands für ein auf den deutschen Kontext zugeschnittenes Modell ist grundsätzlich ein stufenweises Verfahren zu empfehlen: In einem ersten Schritt wären die Übergänge auf dem deutschen Arbeitsmarkt in den Blick zu nehmen, um die Branchen zu identifizieren, in denen Suchfraktionen besonders ausgeprägt sind. Hier könnte auf den Arbeiten von Fitzenberger und Garloff (2007, 2008) aufgebaut werden. Der Fokus würde allerdings stärker auf den für die Mindestlohnevaluation besonders relevanten Niedriglohnsektor gerichtet. Aufbauend auf dieser Datenaufbereitung und -analyse könnte in einem zweiten Schritt ein nach Branchen und möglicherweise Qualifikationsstufen segmentiertes Suchmodell à la van den Berg und Ridder (1998) geschätzt werden. Erst in einem dritten und letzten Schritt würden dann die Spillovereffekte zwischen den Branchen und Qualifikationsstufen modelliert.

Die Alternative zum wage-posting-Ansatz, der in den Studien von Holzner und Launov (2010) und van den Berg und Ridder (1998) verwendet wird, ist ein Bargaining-Modell. Hier erscheint insbesondere die Arbeit von Ahn et al. (2011) interessant, da auch die Möglichkeit positiver Beschäftigungswirkungen des Mindestlohns berücksichtigt ist. Allerdings scheint hier der Aufwand für eine Anpassung an den deutschen Kontext noch einmal höher als beim wage posting-Ansatz, da das Modell von Ahn et al. weder segmentierte Arbeitsmärkte noch eine Substitution zwischen verschiedenen Qualifikationsniveaus vorsieht und auch die übrigen Anwendungen des Bargaining-Ansatzes auf den Kontext der amerikanischen Mindestlöhne zugeschnitten sind. Chowdry et al. (2009, S. 10) geben außerdem grundsätzlich zu bedenken, dass der wage-

posting-Ansatz das tatsächliche Verhalten der Unternehmen möglicherweise besser beschreibt – direkte Verhandlungen über den Lohn zwischen Arbeitnehmern und Arbeitgebern sind vor allem für den Bereich niedriger Einkommen, der ja für die Analysen zum Mindestlohn besonders relevant ist, vermutlich eher die Ausnahme.

Abschließend sei noch auf einige grundsätzliche Grenzen der Suchmodelle im allgemeinen Gleichgewicht für die Evaluation von Mindestlohnwirkungen hingewiesen (vgl. Chowdry et al. 2009, S. 20), die sowohl den wage-posting als auch den Bargaining-Ansatz betreffen.

Zum einen wird in diesen Modellen nur der Arbeitsmarkt modelliert. Mögliche Anpassungen auf den Gütermärkten bleiben außen vor; es kann also nicht untersucht werden, in welchem Umfang die Firmen die durch den Mindestlohn gestiegenen Arbeitskosten über Preiserhöhungen an die Verbraucher weiterreichen. Durch eine solche Anpassung auf den Gütermärkten fielen die Beschäftigungswirkungen vermutlich geringer aus.

Eine zweite Einschränkung der Modelle ist, dass nur abgebildet wird, wie viele Beschäftigungsverhältnisse entstehen oder verloren gehen. Anpassungen beim zeitlichen Umfang oder der Intensität der Tätigkeit sind ebenso wenig erfasst wie der gesamte Komplex der zusätzlichen Lohnbestandteile (Trinkgelder, freiwillige Zusatzleistungen der Firmen), über die in der Realität eine Mindestloohnerhöhung möglicherweise zum Teil kompensiert wird.

Drittens schließlich blenden die Suchmodelle im allgemeinen Gleichgewicht die Substitutionsmöglichkeiten zwischen Kapital und Arbeit weitgehend aus. Eine Reaktion der Unternehmen auf die durch den Mindestlohn teurer gewordene Arbeitskraft könnte aber gerade darin bestehen, dass die Firmen verstärkt auf eine kapitalintensive Produktion setzen. Wie Sorkin (2013) zeigt, benötigen solche Anpassungsreaktionen Zeit, da sich Kapital und Arbeit häufig nicht beliebig schnell substituieren lassen. Die Umstellung der Produktion erfolgt also nicht sofort, sondern in einzelnen, größeren Schritten. Die Firmen werden also vor allem den Kapitaleinsatz intensivieren, wenn sie davon ausgehen, dass Mindestlohnanhebungen von Dauer sind, also nicht durch den Preisanstieg neutralisiert werden. Von daher dürften die kurz- und langfristigen Wirkungen eines Mindestlohns auseinanderfallen. Dies ist ein Problem für viele Studien mit dem DvD-Ansatz, die häufig einen kurzen Zeithorizont be-

trachten. Obwohl ein auf den deutschen Kontext zugeschnittenes Suchmodell im Idealfall um diese Art der Anpassungsreaktion erweitert werden sollte, empfiehlt es sich zunächst, im vorgeschlagenen stufenweisen Verfahren die bestehenden, robusten Modelle an den deutschen Kontext anzupassen. Dadurch lassen sich bereits Erkenntnisse zur Marktmacht und den zu erwartenden Mindestlohnwirkungen gewinnen, bei deren Interpretation natürlich die Grenzen der Modelle mitbedacht werden müssen.

4.2 Berechenbare Allgemeine Gleichgewichtsmodelle

4.2.1 Ziele

Ziel berechenbarer allgemeiner Gleichgewichtsmodelle²⁸ (*Computable General Equilibrium Models*, kurz: *CGE-Modelle*) ist es, gesamtwirtschaftliche Auswirkungen wirtschaftspolitischer Eingriffe zu quantifizieren. Während - ähnlich wie bei strukturellen ökonometrischen Modellen - der Schwerpunkt der Methodik auf der expliziten Formulierung von Verhaltenshypothesen rational handelnder Individuen liegt, nehmen allgemeine Gleichgewichtsmodelle zusätzlich eine gesamtwirtschaftliche Perspektive ein. Dem zugrunde liegt der Gedanke, dass wirtschaftspolitische Eingriffe in einem Wirtschaftsbereich in der Regel nicht nur Anpassungsreaktionen in dem betroffenen Markt bzw. der betroffenen Branche hervorrufen, sondern über Marktinterdependenzen auch indirekte Effekte in sämtlichen anderen Wirtschaftsbereichen verursachen können. Das Ausmaß der indirekten Effekte wird dabei umso höher ausfallen, je größer die Anzahl der betroffenen Akteure ist. In Abhängigkeit von der Richtung und vom Ausmaß der indirekten Effekte können die direkten Effekte wirtschaftspolitischer Maßnahmen entweder gestärkt oder geschwächt werden. CGE-Modelle stellen somit ein Modellinstrumentarium zur Verfügung, das es nicht nur ermöglicht, Nettoeffekte wirtschaftspolitischer Eingriffe zu quantifizieren, sondern mit Hilfe der Formulierung von Verhaltenshypothesen rational handelnder Individuen darüber hinaus die zugrunde liegenden Wirkungsmechanismen zu identifizieren.

²⁸ Bei den folgenden Ausführungen handelt es sich um eine gekürzte und modifizierte Version aus Gürtzgen (2003).

Politiksimulationen mit Hilfe von CGE-Modellen werden üblicherweise durch den Vergleich zweier Gleichgewichtszustände durchgeführt. Durch die Gegenüberstellung eines Referenzszenarios ohne Politikänderung (z.B. eine Ausgangssituation ohne Mindestlohn) mit einem Szenario nach einem wirtschaftspolitischen Eingriff (etwa nach Einführung eines branchenspezifischen Mindestlohnes) lassen sich quantitative Auswirkungen einer Maßnahme schließlich anhand der Änderung der relevanten ökonomischen Variablen ermitteln. Berechenbare allgemeine Gleichgewichtsmodelle stellen somit vorwiegend ein Modellinstrumentarium zur Ex-ante-Evaluation wirtschaftspolitischer Eingriffe dar, das es ermöglicht, Politikanalysen in kontrollierter Umgebung durchzuführen und deren Auswirkungen vor Durchführung der eigentlichen Maßnahme zu ermitteln. Aufgrund ihres Ex-ante-Charakters bieten Simulationen mit Hilfe von CGE-Modellen ebenso wie strukturelle ökonometrische Modelle den Vorteil, einen systematischen Vergleich alternativer Handlungsstrategien vornehmen und potenzielle Zielkonflikte zwischen verschiedenen wirtschaftspolitischen Zielen aufzeigen zu können.

4.2.2 Grundstruktur

Allgemeine Gleichgewichtstheorie: Die allgemeine Gleichgewichtstheorie stellt die wirtschaftstheoretische Fundierung von CGE-Modellen dar. Gegenstand der allgemeinen Gleichgewichtstheorie ist die Untersuchung von Gleichgewichtszuständen auf Grundlage der Formulierung von Verhaltenshypothesen rational handelnder Individuen. Gleichgewichtszustände sind dadurch charakterisiert, dass die individuellen Wirtschaftspläne in Angebot und Nachfrage übereinstimmen und kein Individuum mehr den Anreiz hat, sein Verhalten zu ändern. Unter der Annahme vollkommener Märkte werden die individuellen Entscheidungen über flexible Preise koordiniert, die Nachfrage und Angebot in Übereinstimmung bringen. Ziel der Gleichgewichtstheorie ist es, preisabhängige Angebots- und Nachfragefunktionen für Güter und Produktionsfaktoren abzuleiten und für diese markträumende Preise zu bestimmen. Gleichgewichtszustände sind jedoch nicht zwangsläufig dadurch gekennzeichnet, dass alle relevanten Märkte geräumt sind. Marktunvollkommenheiten in Form von Preisrestriktionen bzw. Mengenrestriktionen können ebenso in allgemeinen Gleichgewichtsmodellen Berücksichtigung finden. Zum Beispiel stellt die Existenz eines Mindestlohnes eine Restriktion in Bezug auf den Preis des Faktors

Arbeit dar. Derartige Marktrigiditäten müssen nicht notwendigerweise exogener Natur sein, sondern können ebenfalls aus dem Optimierungskalkül rational handelnder Wirtschaftssubjekte abgeleitet werden. Angewandt auf den Kontext branchenspezifischer Mindestlöhne, die mittels Allgemeinverbindlicherklärungen von Tarifverträgen eingeführt werden, bedeutet dies, dass Mindestlöhne als Ergebnis nutzenmaximierenden Verhaltens der Tarifparteien endogen im allgemeinen Gleichgewicht bestimmt werden können.

Zentraler Vorteil des allgemeinen Gleichgewichtsansatzes im Vergleich zu den vorher vorgestellten Methoden ist, dass wichtige Wechselwirkungen zwischen Märkten berücksichtigt werden können. Gesamtwirtschaftliche Effekte wirtschaftspolitischer Maßnahmen lassen sich somit unter Berücksichtigung sämtlicher angebots- und nachfrageseitiger Anpassungsreaktionen auf allen Märkten erklären. Dies ermöglicht die Ermittlung von Nettoeffekten von Maßnahmen, die gegebenenfalls erheblich von Partialeffekten abweichen können.

Während beispielsweise in einer Partialanalyse Beschäftigungseffekte von branchenspezifischen Mindestlöhnen lediglich für die betroffene Branche ermittelt werden, berücksichtigt die allgemeine Gleichgewichtsanalyse bei der Ermittlung der Nettoeffekte auch Anpassungsreaktionen in anderen Branchen. Führt z.B. eine mindestlohnbedingte Lohnerhöhung zu Güterpreiserhöhungen in der betroffenen Branche, so kann die Einführung eines branchenspezifischen Mindestlohnes in Abhängigkeit von der Substitutionsbeziehung zwischen Gütern unterschiedlicher Sektoren ebenfalls Beschäftigungseffekte in anderen nur indirekt betroffenen Branchen hervorrufen. Des Weiteren ermöglicht eine allgemeine Gleichgewichtsanalyse unter Berücksichtigung sämtlicher Anpassungsreaktionen mögliche Rückwirkungen auf die Konsumstruktur und das Sparverhalten. Ebenso können fiskalische Effekte einer Mindestlohneinführung ermittelt werden, die z.B. durch ein verändertes Steueraufkommen und veränderte Transferzahlungen an arbeitslose Individuen verursacht werden können.

Modellspezifikation

Akteure: Als elementare Akteure werden in den meisten Modellspezifikationen üblicherweise Haushalte, Unternehmen sowie der Staat unterschieden. Inwiefern das Verhalten weiterer Akteure zu modellieren ist, hängt von der zu untersuchenden Fragestellung ab. Die Modellierung der Einführung von Min-

destlöhnen auf Basis von Allgemeinverbindlicherklärungen erfordert z.B. zwingend die Berücksichtigung von Gewerkschaften als relevante Wirtschaftssubjekte. Hiermit kann die Bildung von Tariflöhnen als Ergebnis nutzen- und gewinnmaximierenden Verhaltens der relevanten Akteure im Modell erfasst werden (siehe nächster Abschnitt).²⁹

Gleichungen: In einem Modell mit Unternehmen, Haushalten/Individuen, dem Staat sowie Gewerkschaften sind Verhaltenshypothesen der zu berücksichtigenden Akteure aufzustellen: Unternehmen maximieren ihren Gewinn bei gegebener Technologie und treffen unter Berücksichtigung der entsprechenden Güter- und Faktorpreise ihre kostenminimierenden Faktornachfrage- bzw. ihre gewinnmaximierenden Güterangebotsentscheidungen. Haushalte bzw. Individuen wählen ihr Faktorangebot und ihre Güternachfrage so, dass die resultierenden Kombinationen aus Konsum und Freizeit bei gegebener Budgetrestriktion ihren Nutzen maximieren. Der Staat stellt öffentliche Güter und Transferleistungen zur Verfügung, die durch die Erhebung von Steuern finanziert werden. Gewerkschaften maximieren ihren Nutzen unter der Restriktion der Arbeitsnachfrage der Unternehmen. Für die Güter- und Faktormärkte, für die vollkommene Märkte angenommen werden, bringt ein flexibler Preismechanismus Angebot und Nachfrage zum Ausgleich. Die Quantifizierung gesamtwirtschaftlicher Effekte wirtschaftspolitischer Eingriffe setzt schließlich eine algebraische Formulierung der oben beschriebenen Verhaltenshypothesen der relevanten Akteure voraus. Konkret lassen sich in CGE-Modellen die erforderlichen Gleichungen in drei Kategorien einteilen (vgl. Klepper et al., 1994):

- Gleichungen, die das Angebots- und Nachfrageverhalten der relevanten Akteure beschreiben,
- Gleichungen, die die Verteilung des Einkommens auf die Akteure bestimmen,

²⁹ Vgl. z.B. Boeters und Savard (2011) für eine Darstellung, wie gewerkschaftliche Lohnverhandlungen in CGE-Modellen berücksichtigt werden können.

- Gleichgewichtsbedingungen für die Güter- und Faktormärkte sowie Identitätsbedingungen für zentrale makroökonomische Aggregate.

Um quantitative Simulationen wirtschaftspolitischer Eingriffe vornehmen zu können, reicht es jedoch nicht aus, die oben genannten Gleichungen in allgemeiner analytischer Form zu spezifizieren. Die numerische Berechenbarkeit des resultierenden Gleichungssystems erfordert vielmehr eine Spezifikation konkreter funktionaler Formen von Nutzenfunktionen und Technologien, um die relevanten Angebots- und Nachfragefunktionen auf Güter- und Faktormärkten bestimmen zu können. Die Zugrundelegung funktionaler Formen kann den Gültigkeitsbereich der mit Hilfe der Simulationen getroffenen Aussagen unter Umständen erheblich eingrenzen. Insbesondere kann die Auswahl bestimmter Funktionen Einschränkungen von Parameterwerten bedingen, die mit empirisch beobachteten Werten nicht in Übereinstimmung zu bringen sind.³⁰ Andererseits erfordert die analytische Handhabbarkeit im Allgemeinen einfache Funktionsspezifikationen, die gewissen Regularitätsanforderungen genügen und eine analytische Bestimmung der Parameterwerte ermöglichen sollten. So ist z.B. die Anzahl der frei bestimmbaren Parameter schon durch die Anzahl der Gleichungen im Modell restringiert. Bei der Auswahl der funktionalen Formen ist daher in der Regel zwischen analytischer Handhabbarkeit und implizierten Einschränkungen des Gültigkeitsbereichs abzuwägen.

Datenbasis und Kalibrierung: Wichtige Voraussetzung für die Durchführung komparativer Politikanalysen mit Hilfe von CGE-Modellen ist, dass die Ausgangssituation ohne Politikeingriff („Benchmark“) als Gleichgewichtslösung des Modells reproduziert werden kann. Das theoretische Modell muss hierzu mit einer mikroökonomisch konsistenten Datenbasis gefüllt werden, die den beobachteten Gleichgewichtszustand der zugrunde liegenden Volkswirtschaft in einem Basisjahr widerspiegelt. Zur Erstellung dieses Datensatzes sind Informationen aus verschiedenen Datenquellen (Input-Output-Tabellen, Statistiken

³⁰ So kann beispielsweise die Auswahl bestimmter funktionaler Formen von Produktionstechnologien zur Folge haben, dass Substitutionsbeziehungen zwischen verschiedenen Produktionsfaktoren nicht variieren können.

zur Einkommensverteilung, Beschäftigungsstatistiken, Einkommen-Verbrauchs-Statistiken, Steuerstatistiken etc.) heranzuziehen.

Damit die in der Benchmarksituation beobachteten Transaktionen mit der konkreten Modellformulierung im Einklang stehen, müssen die den funktionalen Formen zugrunde liegenden Parameter schließlich so bestimmt werden, dass das empirisch beobachtete Referenzgleichgewicht reproduziert wird. Zur Bestimmung der Parameter werden in der Regel zwei verschiedene Verfahren miteinander kombiniert: ökonometrische Schätzungen und die Kalibrierung (vgl. Böhringer und Wiegard 2003). Eine ökonometrische Schätzung sämtlicher Parameter des zugrunde liegenden Modells ist allerdings häufig nicht durchführbar, da für die Schätzungen meist keine ausreichenden Zeitreihendaten zur Verfügung stehen. Dies trifft insbesondere für Datenquellen, wie z.B. Input-Output-Tabellen zu, die lediglich in mehrjährigen Abständen erhoben werden (vgl. Klepper et al. 1996).

Vor diesem Hintergrund wird in der Regel auf das Verfahren der Kalibrierung zurückgegriffen, das verkürzt als „Einpunktschätzung“ charakterisiert werden kann. Die das Ausgangsgleichgewicht beschreibenden Gleichungen werden hierbei nach den zu bestimmenden Parametern aufgelöst, die schließlich unter Einsetzen der endogenen Variablen aus den Daten des Basisjahres berechnet werden können. Steht jedoch nur ein Datensatz für ein einziges Basisjahr zur Verfügung, reichen diese Daten nur dazu aus, um Wertanteilparameter zu bestimmen. Elastizitäten, die Reaktionen auf Preisänderungen widerspiegeln, können nur kalibriert werden, wenn mehrere Beobachtungszeiträume zur Verfügung stehen. Aus diesem Grunde werden Werte für Elastizitäten im Allgemeinen aus exogenen Schätzungen aus der Literatur übernommen. Der wesentliche Nachteil bei der Übernahme exogener Schätzergebnisse liegt darin, dass diese in der Regel aus anderen Modellzusammenhängen abgeleitet werden, die insbesondere nicht den sich aus dem Gleichgewichtsmodell ergebenden Interdependenzen zwischen den einzelnen Gleichungen Rechnung tragen. Ein weiteres Problem bei der Kalibrierung liegt darin, dass - anders als bei ökonometrischen Verfahren - die „Güte“ der Parameterschätzung nicht überprüft werden kann. Das primäre Verfahren zur Überprüfung der Güte des Modells ist der so genannte Konsistenztest, der untersucht, ob das Modell mit den kalibrierten Parameterwerten das Ausgangsgleichgewicht reproduziert. Systematische Sensitivitätsanalysen bezogen auf die Parameterspezifizierung,

die die Modellergebnisse auf ihre qualitative Robustheit hin untersuchen, erweisen sich daher bei allen Politiksimulationen als unabdingbar (Böhringer und Wiegard 2003). Eine zentrale Voraussetzung für die darauf folgende Simulation ist schließlich, dass die Parameter als exogen angenommen werden können und nicht durch die Politikinterventionen beeinflusst werden.

4.2.3 Vorgehensweise und Simulationen

Die generelle Vorgehensweise bei der berechenbaren allgemeinen Gleichgewichtsanalyse wird in Abbildung 3 zusammengefasst. In einem ersten Schritt erfolgt die inhaltliche Konkretisierung der Problemstellung, um hieran anschließend zentrale ökonomische Wirkungszusammenhänge identifizieren zu können, die im Modell berücksichtigt werden sollen. Die Identifikation der ökonomischen Zusammenhänge ist wesentliche Voraussetzung dafür, um numerische Simulationsergebnisse auf ihre Konsistenz hin untersuchen zu können. Die Konstruktion der Datenbasis und die Formulierung des theoretischen Modells erfolgen in der Regel simultan, da Disaggregationsmöglichkeiten sowie die Berücksichtigung bestimmter theoretischer Wirkungszusammenhänge zentral durch die Datenverfügbarkeit beschränkt sind. Die Modellparameter werden schließlich entweder aus der Literatur übernommen oder endogen mit Hilfe des Ausgangsdatensatzes kalibriert. Hiernach erfolgt der so genannte Konsistenztest bzw. „Replication Check“, der überprüft, ob das Modell in der Lage ist, den Ausgangsdatensatz zu reproduzieren. Nach erfolgreichem Replication Check können Simulationsrechnungen für die definierten Politikscenarien durchgeführt werden. Zuletzt erfolgt schließlich die Auswertung der Ergebnisse, die mit Hilfe von Sensitivitätsanalysen bezogen auf zentrale Verhaltensparameter auf ihre Robustheit hin überprüft werden sollten.

4.2.4 Anwendbarkeit auf die Evaluation branchenspezifischer Mindestlöhne

Eine mögliche Anwendung des CGE-Instrumentariums auf die Evaluation branchenspezifischer Mindestlöhne wirft zunächst die Frage nach dem angemessenen Aggregations- bzw. Disaggregationsgrad des Modells auf. Auch wenn die meisten Anwendungen sich auf einem hohen Aggregationsniveau bewegen, sind in CGE-Modellen disaggregierte Analysen prinzipiell für sämtliche im Modell betrachteten Akteure möglich. Die wichtigste Restriktion stellt hierbei die

Datenverfügbarkeit dar. So erfordert die Konstruktion einer mikroökonomisch konsistenten Datenbasis (siehe hierzu Abschnitt 4.2.2) im Rahmen einer sektoral disaggregierten Analyse die Verfügbarkeit von branchenspezifischen Daten, während die Abbildung von Heterogenität im Haushalts- oder Unternehmenssektor die Verfügbarkeit entsprechender Individual bzw. Unternehmensdaten voraussetzt. Soll die Modellspezifikation ebenfalls unterschiedliche Verhaltensfunktionen für heterogene Wirtschaftssubjekte zulassen, sind hierfür zusätzlich disaggregierte Parameterwerte erforderlich.

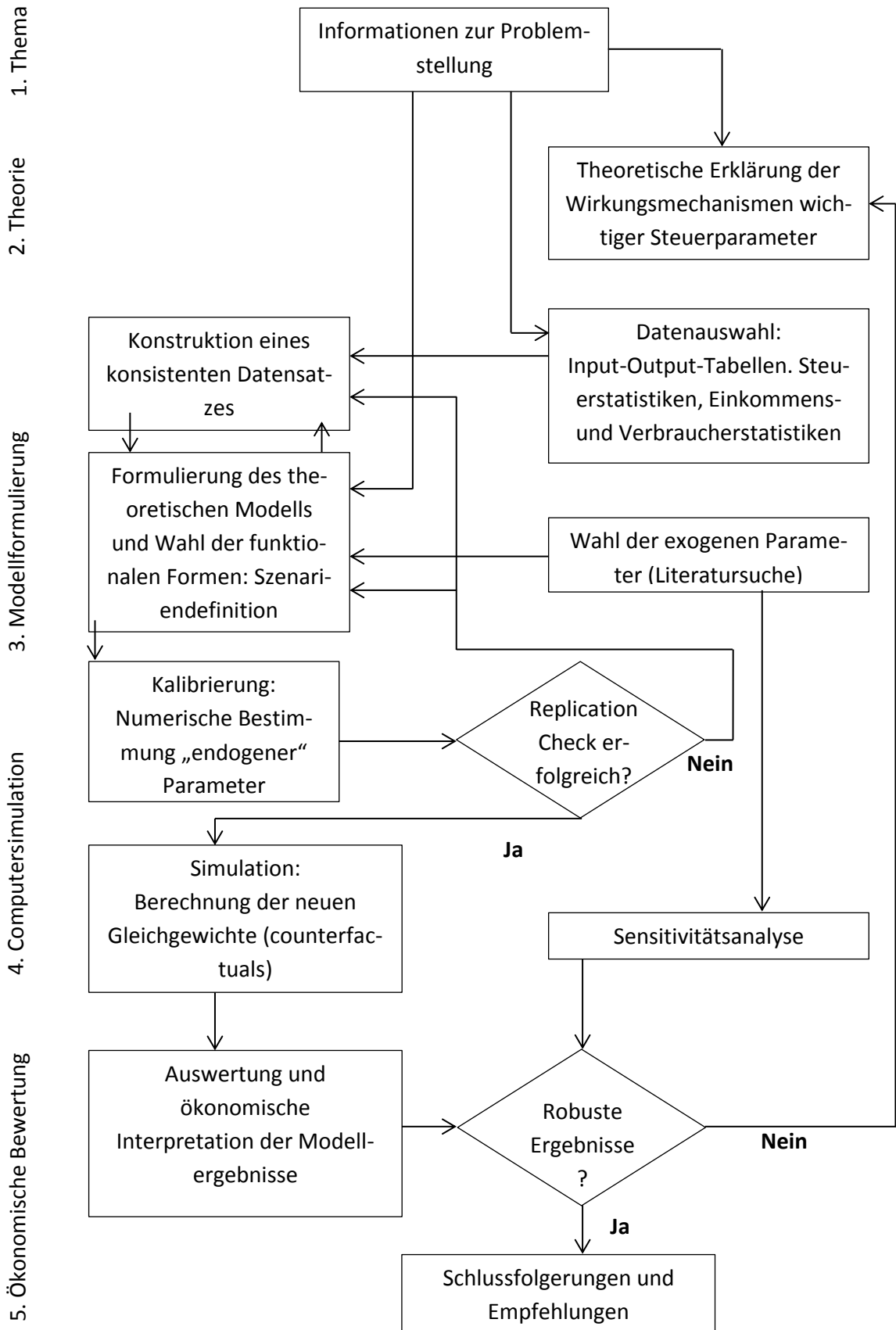
Abgesehen von der Datenverfügbarkeit ist das wichtigste Kriterium für die Auswahl des Aggregationsgrades die der Evaluation zugrunde liegende Fragestellung. So setzt die Einführung von branchenspezifischen Mindestlöhnen in jedem Fall eine sektoral disaggregierte Modellierung voraus. Die Modellierung der Einführung von Mindestlöhnen über Allgemeinverbindlicherklärungen erfordert zudem die Berücksichtigung der Heterogenität der Tarifbindung von Unternehmen. Sollen ebenso distributive Konsequenzen auf der Individual-ebene ermittelt werden, ist ebenfalls eine hinreichende Disaggregation von Haushalten/Individuen nötig. Möchte man weiterhin berücksichtigen, dass Individuen und Unternehmen innerhalb einer Branche nicht nur heterogen im Hinblick auf die Betroffenheit von einer Mindestlohneinführung, sondern auch im Hinblick auf ihre Anpassungsreaktionen sind, bedarf es der Verfügbarkeit entsprechender disaggregierter Reaktionsparameter (wie z.B. Arbeitsnachfrage- und Arbeitsangebotselastizitäten). Um auf Unternehmensebene Substitutionsmöglichkeiten zwischen vom Mindestlohn betroffenen sowie nicht betroffenen Beschäftigten zuzulassen, muss es zudem mindestens einen Qualifikationstyp geben, für den der Mindestlohn bindend ist. Dies erfordert die Modellierung von Arbeitsnachfragefunktionen, die genug Heterogenität im Hinblick auf die nachgefragten Gruppen von Beschäftigten erlauben. In den meisten Anwendungen wird nur zwischen zwei Qualifikationstypen unterschieden, was für eine hinreichende Berücksichtigung der Betroffenheit sowie von Substitutionsmöglichkeiten nicht ausreichen dürfte.

Ein Beispiel für ein CGE-Modell auf Basis deutscher Daten ist das Modell PACE-L, das in mehrjähriger Arbeit am ZEW entwickelt wurde (vgl. Böhringer et al. 2005, Arntz et al. 2008). PACE-L lässt zwar Heterogenität auf Branchen- und Haushaltsebene, jedoch keine Heterogenität auf der Unternehmensseite zu. Ebenso blendet die derzeitige Modellierung des Arbeitsmarktes Marktmacht

von Unternehmen aus. Marktmacht existiert in dem Modell nur auf der Arbeitnehmerseite in Form von Gewerkschaften, deren Lohnverhandlungen mit dem repräsentativen Unternehmen die branchenspezifischen Tariflöhne endogen determinieren. Da die Tarifbindung für den gesamten Sektor (bzw. das repräsentative Unternehmen des Sektors) als exogen angenommen wird, erlaubt das Modell in seiner derzeitigen Form keine Evaluation branchenspezifischer Mindestlöhne. Diese erfordert – wie oben dargelegt – die Berücksichtigung der Heterogenität der Unternehmen, deren Betroffenheit vom Mindestlohn maßgeblich von der Tarifbindung in der Ausgangssituation abhängt. Weiterhin lässt die Modellierung der Arbeitsnachfrage derzeit nur zwei Qualifikationstypen (niedrig und hoch qualifizierte Arbeit) zu. Die Berücksichtigung weiterer Gruppen von Beschäftigten erfordert die ökonometrische Schätzung zumindest sektorspezifischer Arbeitsnachfragefunktionen nach heterogener Arbeit, auf deren Bedeutung für die Abschätzung von Skalen- und Substitutionseffekten schon mehrfach im Kontext der möglichen Verzerrung der DvD-Ergebnisse hingewiesen worden ist. In dem gewünschten Disaggregationsgrad liegen solche Schätzungen jedoch für die vom Mindestlohn betroffenen Branchen derzeit nicht vor.

Eine entsprechende Erweiterung des Modells ist – nicht zuletzt infolge der mangelnden Verfügbarkeit disaggregierter Reaktionsparameter – unserer Einschätzung nach äußerst aufwändig und wäre mit einem erheblichen zeitlichen Vorlauf verbunden. So setzt die zusätzliche Disaggregation auf Unternehmensebene die Unterfütterung des bestehenden Modells mit entsprechenden verknüpften Arbeitgeber-Arbeitnehmerdaten (wie z.B. dem Linked-Employer-Employee Datensatz des IAB) und deren Harmonisierung mit den bestehenden Datenquellen voraus. Weiterhin müssten für den gewünschten sektoralen Disaggregationsgrad strukturelle Arbeitsnachfragemodelle geschätzt werden, um nach Branchen getrennt die relevanten Reaktionsparameter der Arbeitsnachfrage bestimmen zu können. Die Berücksichtigung von Marktmacht wäre besonders aufwändig, da sie zunächst die ökonometrische Bestimmung der Friktionsparameter für die unterschiedlichen Branchen erfordern würde (vgl. hierzu die Darstellung der Methoden und möglichen Datenquellen in Kapitel 4.1). Insgesamt gehen wir davon aus, dass eine sinnvolle Erweiterung des bestehenden Modells eine Bearbeitungszeit von mindestens 24 Monaten erfordern würde.

Abbildung 3: Vorgehensweise bei CGE-Modellen



5 Gesamtfazit für die Erweiterung der Evaluationsmethoden

Ziel dieses abschließenden Kapitels ist es, ein Gesamtfazit im Hinblick auf die Möglichkeiten einer verbesserten und erweiterten Evaluation der Einführung branchenspezifischer Mindestlöhne zu ziehen. Anders als die in Kapitel 3 und 4 abgegebenen allgemeinen Bewertungen werden hier die in Kapitel 2 herausgearbeiteten unterschiedlichen Ausgangslagen der verschiedenen Branchen vor allem im Hinblick auf die Datengrundlage und die institutionellen Rahmenbedingungen berücksichtigt.

Unabhängig vom konkreten Branchentyp haben die vorangegangenen Ausführungen gezeigt, dass die mögliche Endogenität der Einführung bzw. Erhöhung eines Mindestlohns eine zentrale und schwer lösbare Herausforderung darstellt. Wenn die Einführung eines Mindestlohns bzw. dessen Veränderung eine endogene Reaktion auf die vergangene und aktuelle Branchensituation darstellt, erschwert dies das Auffinden einer Kontrollgruppe, welche die kontrafaktische Situation der vom Mindestlohn betroffenen Branche abbilden kann. Grundsätzlich lassen sich die Möglichkeiten, dieses Problem zu reduzieren, in zwei Gruppen einteilen.

Zum einen lässt sich das Problem der Politikendogenität umgehen, indem die Variation der Eingriffsintensität innerhalb der betroffenen Branche zur Identifikation der Mindestlohnwirkungen genutzt werden. In diese Kategorie fallen unter anderem die in der ersten Evaluationsrunde gewählten DvD-Ansätze, die Betriebe oder Beschäftigte mit einem unterschiedlichen Ausmaß der Betroffenheit miteinander vergleichen. Allerdings laufen diese Ansätze eher Gefahr, durch Spillovereffekte verzerrt zu sein. Eine mögliche Weiterentwicklung dieses Ansatzes besteht daher in der in Kapitel 3.1.3 diskutierten Verwendung regional variierender Eingriffsintensitäten, bei der zumindest für einige Branchen mit geringeren Spillovereffekten zu rechnen ist. Allerdings hat die Diskussion auch gezeigt, dass dieser Ansatz nur für die Evaluation der Einführung eines Mindestlohns, nicht jedoch für die Evaluation sukzessiver Erhöhungen geeignet ist.

Ein alternativer Ansatz besteht darin, eine Kontrollgruppe außerhalb der betroffenen Branche zu verwenden und die Endogenität der Mindestlohneinfüh-

rung dadurch abzumildern, dass eine größtmögliche Vergleichbarkeit von Treatment- und Kontrollgruppe auf Branchenebene hergestellt wird. In diese Kategorie fällt die in diesem Bericht diskutierte Methode der synthetischen Kontrollgruppe. Diese stellt zumindest für die Zeit vor der Mindestlohneinführung eine Vergleichbarkeit auf Branchenebene her und gewährleistet damit, dass die Common Trends Annahme eher gültig ist als im Falle eines Verfahrens, welches lediglich auf die Vergleichbarkeit von Beschäftigten oder Betrieben abzielt. Grundsätzlich gilt jedoch, dass dieser Ansatz das Problem der Politikendogenität nicht löst, sondern lediglich abschwächen kann.

Da jeder der Ansätze spezifische Stärken und Schwächen aufweist, sollte eine zukünftige Evaluation die verschiedenen Ansätze kontrastieren und die Natur der gefundenen Abweichungen unter Berücksichtigung der zugrundeliegenden Annahmen interpretieren. Gerade die Abweichungen unter Verwendung verschiedener Kontrollgruppen könnten per se Anhaltspunkte für die ökonomische Wirkungsweise der Mindestlöhne liefern (beispielsweise weil die verschiedenen Kontrollgruppen von unterschiedlichen Spillovereffekten betroffen sein könnten, die explizit zu untersuchen wären). Hierbei ist auch das Zusammenspiel der Effekte für verschiedene Ergebnisvariablen vor dem Hintergrund empirisch plausibler Verhaltensreaktionen (Arbeitsnachfrageelastizitäten) zu interpretieren. Die Verzerrung der geschätzten DvD-Effekte aufgrund von Skalen- und Substitutionseffekten könnten mit Hilfe eines strukturellen Arbeitsnachfragemodells erfasst werden, wenn auf dem Arbeitsmarkt keine monopolistische Konkurrenz vorliegt.

Vor diesem Hintergrund ist davon auszugehen, dass einfache DvD-Ansätze mit auf ex-ante Vorüberlegungen basierenden Kontrollgruppen außerhalb und innerhalb der betroffenen Branche weiterhin Teil zukünftiger Evaluationsstudien sein werden. Darüber hinaus lassen sich jedoch für alle Branchen Methoden identifizieren, die im Vergleich zu diesen Ansätzen die Möglichkeit einer methodischen Verbesserung bieten. Im Folgenden werden die methodischen Ansätze daher unter Berücksichtigung der zukünftigen Datenlage bewertet und nach ihrem Verbesserungspotenzial geordnet. Insbesondere lässt sich für jede Branche eine „Best Practice“ identifizieren. Allerdings kann auch eine „Best Practice“ zumeist nicht alle Schwierigkeiten einer kausalen Identifikation überwinden, so dass die „Best Practice“ immer auch den Vergleich mit metho-

dischen Ansätzen vorsehen sollte, die gerade auf anderen identifizierenden Annahmen beruhen.

Die Bewertung wird im Folgenden zur besseren Strukturierung getrennt nach der in der Einleitung vorgenommenen Typisierung der Branchen vorgenommen, welche die drei folgenden Gruppen umfasst: 1) traditionelle Handwerke mit Facharbeiterdominanz, 2) einfache, standardisierte Dienstleistungen und 3) Branchen mit privaten und öffentlichen Anbietern und einer spezifischen Wettbewerbssituation auf dem Gütermarkt. Wie die folgenden Ausführungen jedoch zeigen werden, schließt die Tatsache, dass die drei Branchentypen strukturell ähnliche Branchen umfassen, nicht grundsätzlich aus, dass sich innerhalb der Branchentypen differenzierte Bewertungen ergeben können.

5.1 Branchentyp I (Bau und Baunebengewerbe)

Zukünftige Verbesserung der Datenlage: Die Branchen des Bauhaupt- und Baunebengewerbes verfügen im Vergleich zu den anderen Branchen mit einem gesetzlichen Mindestlohn über die beste Datenlage. Aufgrund der bereits langen Laufzeiten des gesetzlichen Mindestlohns stehen die administrativen Daten der BA für die Analysen zur Verfügung und wurden in der ersten Evaluationsrunde bereits umfassend genutzt. Die dabei entstehenden Ungenauigkeiten des betrieblichen Geltungsbereichs des Mindestlohns sind damit jedoch nicht reduzierbar.

Zur verbesserten Abgrenzung der persönlichen Betroffenheit in den IEB-Daten müssen Validierungsdatensätze zur Verfügung stehen, die eine möglichst präzise Imputation der fehlenden Arbeitszeitinformaton bzw. der fehlende Stundenlohnangabe ermöglichen. Für die meisten Branchen des Typs I steht dazu primär der Mikrozensus zur Verfügung, der jedoch nur eine grobe Imputation von durchschnittlichen tatsächlichen Arbeitszeiten für Teil- und Vollzeitbeschäftigte erlaubt. Die Präzision einer solchen Durchschnittsimputation hängt insbesondere von der Heterogenität der Arbeitszeiten in der jeweiligen Branche ab. Für den typischen Vollzeitbeschäftigten in den Branchen des Typs I ist vor allem mit einer betriebsspezifischen Variation der Arbeitszeiten zu rechnen, der jedoch eventuell durch entsprechend verfeinerte Durchschnittswerte Rechnung getragen werden kann. Für Teilzeitbeschäftigte ist hingegen von einer größeren Heterogenität auf individueller Ebene auszugehen, so dass die

Durchschnittsimputation hier an ihre Grenzen stößt. Da die Baubranche jedoch nur in geringem Maße Teilzeitbeschäftigte einsetzt, lässt sich mit einer Durchschnittsimputation ein vermutlich zufrieden stellendes Ergebnis erzielen. Eine präzisere Imputation konnte in der ersten Evaluationsrunde lediglich im Dachdeckerhandwerk unter Verwendung der Daten der Lohnausgleichskasse des Dachdeckerhandwerks umgesetzt werden. Ein ähnliches Imputationsverfahren wäre zukünftig auch unter Verwendung der Daten aus dem Meldeverfahren der Urlaubskasse des Bauhauptgewerbes (SOKA Bau) möglich. Darüber hinaus ließe sich für die übrigen Branchen ein Validierungsdatensatz erheben, der die fehlenden Informationen für eine Stichprobe von Unternehmen aus der Branche erhebt. Dies ermöglicht bei ausreichender Fallzahl eine Imputation der fehlenden Arbeitszeiten und kann zudem zur Abschätzung des Messfehlers herangezogen werden, wie dies in der ersten Evaluationsrunde bereits für das Bauhauptgewerbe erfolgt ist. Diese Vorgehensweise ließe sich auch auf andere Baunebenbranchen übertragen.

Ansätze mit keinem oder nur geringem Verbesserungspotenzial sind insbesondere die Diskontinuitätsansätze, Ansätze unter Verwendung von Instrumentvariablen sowie DvD in Kombination mit Matching. **DvD mit Matching** stellt wie in Abschnitt 3.1.1 diskutiert kaum eine Verbesserung gegenüber den bereits angewendeten konditionalen DvD-Schätzern dar. In beiden Fällen wird für beobachtete Unterschiede und zeitinvariante, unbeobachtete Faktoren kontrolliert, einer möglichen Politikendogenität, die sich in unterschiedlichen Branchenentwicklungen niederschlägt, wird dabei jedoch ebenso wenig Rechnung getragen wie in den bereits verwendeten Ansätzen. Die **Diskontinuitätsansätze**, die sich theoretisch die Diskontinuität in der Mindestlohnhöhe entlang der innerdeutschen Grenze zunutze machen könnten, scheiden auf den zweiten Blick vollständig aus. Der mögliche Vorteil der Vergleichbarkeit der Marktsituationen in Grenznähe schlägt in einen Nachteil um, wenn die damit einhergehende Marktintegration zu starken Wechselwirkungen beiderseits der Grenze führt. Darüber hinaus richtet sich der zu zahlende Mindestlohn am Ort der Leistungserbringung und somit im Falle des Bauhaupt- und Baunebengewerbes am Standort des Kunden aus. Die gesetzlich vorhandene und für die Anwendbarkeit der Methode erforderliche Diskontinuität wird somit weitgehend nivelliert. Für die Umsetzung eines **Instrumentenvariablen**schätzers fehlen geeignete Variablen, welche zwar exogen die individuelle Betroffenheit

von einem Mindestlohn beeinflussen, nicht jedoch die Zielgrößen der Evaluation.

Ansätze mit beschränktem Verbesserungspotenzial sind regionale Panelanalysen bzw. inkrementelle DvD-Schätzer sowie die Schätzung struktureller Suchmodelle. Bei **regionalen Panelmodellen** beschränkt allerdings die Endogenität der Eingriffsintensität klar den Nutzen des Ansatzes (Abschnitt 3.1.3). Während der Endogenität regional variierender Mindestlohnhöhen (wie z.B. in den USA) durch eine zusätzliche Instrumentierung der Mindestlohnhöhen durch das regionale Parteienspektrum begegnet werden könnte, lässt sich die regional variierende Eingriffsintensität im Fall einer regional invariablen Mindestlohnhöhe kaum instrumentieren. Dies ist auch der Grund, warum der inkrementelle DvD-Schätzer tendenziell kritisch zu bewerten ist, da der Endogenität der Eingriffsintensitäten über die Zeit nicht Rechnung getragen wird. Trotz dieser Einschränkungen kann die Verwendung der regionalen Variation in der Eingriffsintensität die bisherigen DvD-Ansätze für den Verlauf zwischen zwei Zeitpunkten nur unter Berücksichtigung der Eingriffsintensität im früheren Zeitpunkt (siehe Abschnitt 2.1) auf Betriebs- und Individualebene ergänzen. So leidet der Vergleich von Betrieben mit einer unterschiedlichen Eingriffsintensität im Kontext eines DvD-Schätzers unter ähnlichen Endogenitätsproblemen, ist darüber hinaus jedoch eher von Wechselwirkungen zwischen den Vergleichsbetrieben betroffen. Demgegenüber fallen solche Wechselwirkungen zwischen regionalen Märkten schwächer aus, wenn die Märkte stark regional organisiert sind. Gerade in handwerksnahen Baubranchen ist dies teilweise gegeben. Durch eine sinnvolle Abgrenzung der betrachteten Regionen kann somit die im Betriebskontext problematische SUTVA-Annahme eher eingehalten werden. Die grundlegende Problematik der möglichen Endogenität der Eingriffsintensität bleibt in diesem Ansatz jedoch bestehen.

Strukturelle Suchmodelle sind, wie in Abschnitt 4 diskutiert, grundsätzlich eine sinnvolle Ergänzung der quasi-experimentellen Ansätze. Dies gilt auch im Kontext der Baubranchen. Die Vorteile des Ansatzes – ex-ante Modellierung, Berücksichtigung der Friktionen am Markt, Abschätzung der Marktmacht der Unternehmen – wiegen die Nachteile – Vernachlässigung des Gütermarktes, fehlende Anpassungsmechanismen über Arbeitszeit und zusätzliche Lohnbestandteile, fehlende Substitution zwischen Kapital und Arbeit – angesichts der hohen Entwicklungskosten jedoch nur bedingt auf. Denn gerade in den Bau-

branchen spielen flexible Arbeitszeiten und zusätzliche Lohnbestandteile eine wichtige Rolle. Zudem dürfte die Möglichkeit der Weitergabe von Kostenerhöhungen an die Kunden zumindest in denjenigen Baubranchen mit geringen Substitutionsmöglichkeiten durch Schwarz- oder Eigenarbeit ebenfalls bedeutsam sein. Darüber hinaus hat sich im Dachdeckerhandwerk gezeigt, dass die Substitution zwischen Kapital und Arbeit ebenfalls einen relevanten Anpassungsmechanismus an mindestlohnbedingte Kostenerhöhungen sein kann. Gleichzeitig dürfte die Marktmacht der Arbeitgeber in den Baubranchen schwächer ausgeprägt sein als in Branchentyp II und III. Das Potenzial struktureller Suchmodelle ist somit zwar gegeben, aber aufgrund der spezifischen Branchensituation als geringer einzuschätzen als in den Branchen des Typs II und III.

Die **Best Practice** zur Evaluation der Mindestlohnwirkungen im Bauhaupt- und Baunebengewerbe für die Analyse aggregierter Branchenindikatoren läge in der Verwendung **synthetischer Kontrollgruppen**. Der Ansatz weist den Vorteil auf, dass zur Identifizierung einer geeigneten Kontrollgruppe die Variation über alle möglichen Kontrolleinheiten genutzt wird, anstatt wie bisher alle Kontrollgruppen im Rahmen der Prüfung ihrer Validität gleich zu gewichten. Da es für die einem Mindestlohn unterliegenden Branchen des Bauhaupt- und Baunebengewerbes zahlreiche baunahe Branchen gibt, die bislang keinen Mindestlohn eingeführt haben, ist es für Branchentyp I durchaus realistisch einen „Pool“ aus strukturell ähnlichen Branchen zu finden, die es erlauben durch eine entsprechende Gewichtung eine passgenaue synthetische Kontrollgruppe zu generieren. Die Wirkungen des Mindestlohns auf individuelle Ergebnisgrößen lassen sich mittels dieses Ansatzes jedoch nicht ermitteln. Aus diesem Grund stellt das **CiC-Verfahren**, welches auf Individualebene anwendbar ist, eine sinnvolle Ergänzung des Methodenspektrums dar. Für die Baubranchen ist zudem die notwendige, umfangreiche Datenbasis vorhanden, um diesen Ansatz umsetzen zu können. Vorteilhaft im Vergleich zum DvD-Ansatz ist vor allem die Abschätzung heterogener Mindestlohnwirkungen entlang der Verteilung unterschiedlich skalierteter Ergebnisvariablen. Das Verfahren setzt somit weniger Annahmen voraus als der DvD-Schätzer. Allerdings setzt der Schätzer voraus, dass die Veränderungen in der Verteilung der Ergebnisvariablen in der Kontrollgruppe die kontrafaktischen Veränderungen entlang der Verteilung für die Treatmentgruppe ohne Treatment abbilden. Diese Variante

der Common Trends Annahme ist im Einzelfall kritisch zu hinterfragen. Das Hinzufügen weiterer Kovariate bei der Modellierung der kontrafaktischen Verteilung kann die Plausibilität dieser Annahme sicher erhöhen. Abschließend ist anzumerken, dass sowohl der synthetische Kontrollgruppenansatz als auch der CiC-Ansatz ebenso wie der DvD-Ansatz auf der SUTVA-Annahme beruhen, die Spillovereffekte ausschließt. Eine umfassende Untersuchung möglicher Verzerrungen in Abhängigkeit von den Kontrollgruppen (siehe oben) erweist sich somit auch für diese Ansätze als unabdingbar.

5.2 Branchentyp II (Arbeitsintensive Sektoren)

Zukünftige Verbesserung der Datenlage: Da bei den Wäschereidienstleistungen die Einführung eines Mindestlohnes erst Ende 2009 erfolgte, waren in der ersten Evaluationsrunde amtliche Daten für den Zeitraum nach der Mindestlohnänderung nur beschränkt verfügbar. Für eine zukünftige Evaluation sollten hingegen amtliche Daten mindestens bis zum Jahr 2012 zur Verfügung stehen. Die Probleme bei der Abgrenzung der Betroffenheit der Betriebe bestehen jedoch weiterhin.

Ansätze mit keinem oder nur geringem Verbesserungspotenzial sind ähnlich wie bei den Branchen des Typs I die Diskontinuitätsansätze, Ansätze unter Verwendung von Instrumentenvariablen sowie DvD in Kombination mit Matching. Für die Gebäudereinigung gilt ebenso wie für die Branchen des Typs I, dass die Identifikationsprobleme aufgrund der Möglichkeiten grenzüberschreitender Leistungen, die im Rahmen der Diskontinuitätsansätze auftreten, besonders problematisch sind. Für die Wäschereidienstleistungen hingegen sollten diese Identifikationsprobleme weniger schwer wiegen, da es sich um ortsgebundene Leistungserbringungen handelt.

Ein Ansatz mit lediglich **beschränktem Verbesserungspotenzial** ist für den Branchentyp II auf Basis der vorherigen Überlegungen ebenfalls der **inkrementelle DvD-Ansatz**.

Als **Best Practice** zur Evaluation der Mindestlohnwirkungen in den arbeitsintensiven Branchen stellt sich ebenso wie für die Branchen des Typs I die Verwendung **synthetischer Kontrollgruppen** sowie das **CiC-Verfahren** dar. Sofern man die Treatmentgruppe über die betriebliche bzw. individuelle Betroffenheit vom Mindestlohn abgrenzen möchte, ist genau wie beim DvD-Ansatz die

Voraussetzung für das CiC-Verfahren eine möglichst trennscharfe Abgrenzung des betrieblichen und personellen Geltungsbereichs. Während eine solche Abgrenzung für die Gebäudereinigung auf zumindest auf betrieblicher Ebene gewährleistet ist, ergeben sich für die Wäschereidienstleistungen hier Identifikationsprobleme der vom Mindestlohn betroffenen Betriebe und Beschäftigten. Für diese Branche kann das Verfahren aber zumindest für die Evaluation der Effekte auf aggregierter Branchenebene eingesetzt werden.

Die Anwendung des **synthetischen Kontrollgruppenansatzes** ist in einer zweiten Evaluationsrunde für beide Branchen möglich, da die zukünftige Datenlage die Verfügbarkeit von (administrativen) Daten vor und nach der Intervention gewährleistet. Zum anderen sind beide Branchen über die 5-Steller-Ebene der Wirtschaftszweigklassifikation WZ08 identifizierbar, was die Ermittlung aggregierter Effekte auf der Interventionsebene der Branche ermöglicht. Bei den Wäschereidienstleistungen umfasst die 5-Steller-Klassifikation zwar auch Betriebe, die nicht unter die Mindestlohnregelungen fallen, dies schließt jedoch die Anwendung des synthetischen Kontrollgruppenansatzes nicht grundsätzlich aus. Unter der Voraussetzung, dass es keine Spillovereffekte zwischen betroffenen und nicht betroffenen Betrieben gibt, müssten die auf Basis einer erweiterten Branchenklassifikation ermittelten betragsmäßigen Effekte auf die Ergebnisvariablen als untere Grenze interpretiert werden.

Im Gegensatz zu den Branchen des Typs I sehen wir bei **strukturellen Suchmodellen**, die die Existenz von Suchfraktionen und somit Marktmacht auf Seiten der Arbeitgeber zulassen, ein größeres Verbesserungspotenzial. Zwar gilt auch hier, dass mit den derzeit vorhandenen Modellen wichtige Anpassungsprozesse insbesondere im Hinblick auf Ausweichreaktionen – wie die Substitution regulärer durch geringfügige Beschäftigungsverhältnisse – nicht erfasst werden können. Anders als bei den Branchen des Typs I spricht für die arbeitsintensiven Branchen vor allem der hohe Anteil an Frauen unter den Beschäftigten dafür, dass die zugrundeliegenden Arbeitsmärkte durch mögliche Friktionen und somit Marktmarkt geprägt sein könnten. Der Grund hierfür ist, dass Frauen typischerweise eine geringe Mobilität auf dem Arbeitsmarkt aufweisen. Dies kann zwar durch die Verfügbarkeit von Arbeitsplätzen in anderen gering qualifizierten Branchen abgemildert werden – letztendlich bedarf die Frage nach der Relevanz von Suchfraktionen jedoch empirischer Klärung. Mit den administrativen Individualdaten der BA (IEB), die präzise Informationen zu

verschiedenen Übergängen am Arbeitsmarkt liefern, liegen hier – abgesehen von den oben erwähnten Abgrenzungsproblemen – Daten vor, die für die Quantifizierung der empirischen Bedeutung von Suchfraktionen gut geeignet sind.

5.3 Branchentyp III (Sektoren mit spezifischer Wettbewerbssituation auf dem Gütermarkt)

Zukünftige Verbesserung der Datenlage: Da in dem Pflegesektor und in der Abfallwirtschaft die Einführung eines Mindestlohnes erst 2010 erfolgte, waren in der ersten Evaluationsrunde in beiden Branchen amtliche Daten für den Zeitraum nach der Mindestlohnänderung nicht oder nur beschränkt verfügbar. Für eine zukünftige Evaluation sollten hingegen amtliche Daten mindestens bis zum Jahr 2012 zur Verfügung stehen, so dass zukünftige Analysen sich nicht mehr auf die Unternehmensebene beschränken müssen und ebenfalls auf Individualebene durchgeführt werden können.

Ansätze mit keinem oder nur geringem Verbesserungspotenzial sind ebenso wie bei den Branchen des Typs I und II die Diskontinuitätsansätze, Ansätze unter Verwendung von Instrumentenvariablen sowie DvD in Kombination mit Matching. Für die Diskontinuitätsansätze dürften die Identifikationsprobleme für den Pflegesektor und die Abfallwirtschaft jedoch weniger schwer wiegen. Grundsätzlich resultieren diese daraus, dass grenzüberschreitende Leistungen unter Umständen zu einer vollständigen Marktintegration führen, die Diskontinuitäten im Treatment entlang der innerdeutschen Grenze eliminieren und mit Spillovereffekten einhergehen. Für den Pflegesektor gilt indes, dass der Ort der Erbringung von Pflegeleistungen zumindest bei stationären Einrichtungen mit dem Ort der Einrichtung übereinstimmen sollte. Bei ambulanten Pflegediensten sind grenzüberschreitende Erbringungen von Leistungen zwar grundsätzlich möglich, hier wäre jedoch zu prüfen, welcher Mindestlohn in einem solchen Fall zur Anwendung käme. Die Abfallwirtschaft ist durch eine noch stärkere räumliche Segmentierung des Gütermarktes charakterisiert. Die Marktsegmentierung resultiert in erster Linie aus dem Grundsatz der im Kreislaufwirtschafts- und Abfallgesetz festgelegten Entsorgungsnähe und darauf Bezug nehmende verbindliche Anlagenzuweisungen in den Abfallwirtschaftsplänen, die eine Beseitigung von Abfällen innerhalb des Sammelgebiets vorschreiben. Als weiterer Grund für eine räumliche Marktsegmentierung sind

zudem hohe Transportkostenanteile an den Kosten der Abfallsammlung zu nennen. Allerdings gilt hier ein bundesweit einheitlicher Mindestlohn. Voraussetzung für regionale Diskontinuitäten in der Eingriffsintensität sind demnach unterschiedliche Entlohnungsniveaus vor der Mindestlohneinführung. Ob diese Voraussetzung erfüllt ist, hängt davon ab, inwiefern die Segmentierung regionaler Gütermärkte mit einer Segmentierung der regionalen Arbeitsmärkte einhergeht.

Für den **inkrementellen DvD-Ansatz** bedeutet das relative hohe Ausmaß an räumlicher Marktsegmentierung in der Abfallwirtschaft, dass die mit diesem Ansatz verbundenen **beschränkten Verbesserungen** für die Abfallwirtschaft unter Umständen größer ausfallen als für die anderen Branchen. Ein weiterer Ansatz mit **beschränktem Verbesserungspotenzial** ist – zumindest für die Abfallwirtschaft – die Schätzung **strukturelle Suchmodelle**. Der zentrale Grund für das lediglich eingeschränkte Verbesserungspotenzial liegt ebenso wie bei den Branchen des Typs I darin, dass strukturelle Suchmodelle wichtige Aspekte vernachlässigen, die vor allem für die erwarteten Beschäftigungseffekte von Bedeutung sein können. So haben die Ausführungen in Abschnitt 4.1.8 gezeigt, dass über die Ausblendung der Güternachfrage keine Anpassungen auf den Gütermärkten berücksichtigt werden. Konkret kann also nicht untersucht werden in welchem Umfang Unternehmen die durch den Mindestlohn gestiegenen Arbeitskosten über Preiserhöhungen an die Verbraucher weiterreichen. Dieser Aspekt ist aber gerade für die Abfallwirtschaft von zentraler Bedeutung, bei der in Teilen der Branche das im Gebührenrecht verankerte Kostendeckungsprinzip unmittelbar die Weitergabe erhöhter Lohnkosten über höhere Gebühren an die Endverbraucher ermöglicht. Zudem deuten die Ergebnisse der ersten Evaluationsrunde darauf hin, dass monopsonistische Strukturen auf dem Arbeitsmarkt der Abfallwirtschaft von geringerer Bedeutung sind.

Als **Best Practice** zur Evaluation der Mindestlohnwirkungen in den Branchen des Typs III sehen wir die Verwendung **synthetischer Kontrollgruppen** sowie des **CiC-Verfahrens**, welches ebenfalls in den Branchen des Typs III dazu dienen kann, die Sensitivität der DvD-Schätzungen hinsichtlich der zugrunde gelegten Skalierung zu überprüfen. Zur Anwendung **synthetischer Kontrollgruppen** ist anzumerken, dass diese zwar ein erhebliches Verbesserungspotenzial darstellen können, gleichzeitig aber grundlegende Probleme bei der Umsetzung auftauchen können. Diese resultieren weniger aus der Datenverfügbar-

keit, da die Datenlage in einer zukünftigen Evaluation die Anwendung dieses Ansatzes, welcher Daten vor und nach der Intervention erfordert, grundsätzlich ermöglicht. Weitaus problematischer ist möglicherweise die Konstruktion einer synthetischen Kontrollgruppe. So handelt es sich bei dem Pflegesektor und der Abfallwirtschaft um Branchen, für die es anders als bei den Branchen des Typs I und II in der ersten Evaluationsrunde schwierig war, nur eine einzelne strukturell vergleichbare andere Branche zu identifizieren. Zwar basiert die Idee einer synthetischen Kontrollgruppe darauf, gerade die Variation der Ergebnisvariablen in den potenziellen Vergleichsbranchen auszunutzen, um durch eine entsprechende Gewichtung einen der Treatmentbranche möglichst ähnlichen Verlauf in den jeweiligen Ergebnisvariablen vor der Intervention herzustellen. Unterscheiden sich die Arbeits- und Gütermarktstrukturen in der Treatment- und den Vergleichsbranchen aber fundamental voneinander, kann es Konstellationen geben, bei denen eine spezifische synthetische Kontrollgruppe einen gleichen Verlauf in den Ergebnisvariablen immer nur für einen der beiden Märkte herstellen kann.

Dies lässt sich anhand des folgenden Beispiels veranschaulichen: In der Treatmentbranche werde bei steigenden Löhnen und gleichzeitig steigenden Güterpreisen ein konstanter Beschäftigungsverlauf infolge einer preisunelastischen Güternachfrage beobachtet. In diesem Fall ließe sich ein konstanter Beschäftigungsverlauf prinzipiell durch die entsprechende Gewichtung zweier Vergleichsbranchen mit jeweils steigender und sinkender Beschäftigung herstellen. Hier könnte es sich etwa um zwei Branchen mit vollkommen preiselastischer Güternachfrage und somit konstanten Güterpreisen handeln, von denen eine einen monopsonistischen Arbeitsmarkt und die andere einen kompetitiven Arbeitsmarkt aufweist. In solch einem Fall ließen sich mit der für den Arbeitsmarkt konstruierten synthetischen Kontrollgruppe nicht gleichzeitig die relevanten Gütermarktvariablen der Treatmentbranche, wie Unternehmensgewinne und Güterpreise, reproduzieren. Möchte man Interaktionen von Arbeitsmarkt- und Gütermarktergebnissen berücksichtigen, wäre der synthetische Kontrollgruppenansatz für den Fall, dass sich keine strukturell ähnlichen Vergleichsbranchen finden ließen, somit nicht umsetzbar. Die Tragfähigkeit der Verwendung synthetischer Kontrollgruppen wäre gegebenenfalls empirisch genau zu untersuchen.

Schließlich schätzen wir im Gegensatz zur Abfallwirtschaft das Potenzial **struktureller Suchmodelle**, die Aufschluss über die mögliche Relevanz von Suchfraktionen liefern, für den Pflegesektor größer ein. Zwar wurde in der Studie zum Pflegesektor argumentiert, dass der Arbeitsmarkt für Pflegefachkräfte angesichts des weit verbreiteten Fachkräftemangels eher ein Angebots- als ein Nachfragemarkt ist, der hohe Anteil an Frauen unter den Beschäftigten – ähnlich wie in den Branchen des Typs II – könnte dennoch ein Indiz für mögliche Friktionen und somit Marktmarkt liefern. Auch hier ist die empirische Relevanz von Suchfraktionen letztendlich eine empirische Frage, zu deren Klärung die Schätzung ökonometrischer Suchmodelle einen erheblichen Beitrag leisten kann.

Zur abschließenden Übersicht fasst Tabelle 5 die Bewertung der methodischen Ansätze auf Basis der vorangegangenen Überlegungen für die drei Typen von Branchen zusammen. Die Tabelle zeigt, dass die Bewertungen aufgrund der Heterogenität der Branchen nicht nur über die drei Branchentypen hinweg, sondern auch innerhalb der Typen variieren können. Dies setzt der Übertragbarkeit der Bewertungen auf neue, weitere Mindestlohnbranchen gewisse Grenzen und verdeutlicht, dass die Bewertungen im Einzelfall auf Basis der jeweiligen institutionellen Rahmenbedingungen erfolgen müssen. So haben die vorangegangenen Ausführungen gezeigt, dass die für die Anwendbarkeit des RD-Ansatzes maßgebliche Relevanz des Einsatzortsprinzips auch innerhalb der Gruppen variieren kann. Dies bedeutet in der Konsequenz, dass die Anwendbarkeit für neue Branchen jeweils separat untersucht werden muss. Gleiches gilt für die Bewertung des synthetischen Kontrollgruppenansatzes, sofern einzelne der neu hinzugekommen Branchen durch sehr spezifische Gütermarktbedingungen charakterisiert sind. Bei den strukturellen Suchmodellen ist ebenfalls im Einzelfall zu prüfen, inwiefern die Vorteile des Ansatzes mögliche Nachteile aufwiegen.

Insgesamt verdeutlicht Tabelle 5, dass sich für alle der bisher acht evaluierten Branchen Methoden methodische Erweiterungen identifizieren lassen. Dabei schneiden insbesondere das CiC-Verfahren und der synthetische Kontrollgruppenansatz gut ab und bieten für die meisten Branchen klare Verbesserungspotenziale gegenüber dem bislang in der ersten Evaluationsrunde verfolgten DvD-Ansatz. Darüber hinaus können strukturelle Methoden für einen großen

Teil der Branchen einen wertvollen Beitrag zu einer verbesserten Evaluation der Mindestlohnwirkungen liefern.

Tabelle 5: Bewertung der Ansätze nach Branchentyp

Branchentyp	I	II	III		
Branche	Bauhaupt- und Baunebengewerbe	Gebäude- reinigung	Wäscherei	Pflegesektor	Abfall- wirtschaft
Nicht-Strukturelle/Ex-Post-Methoden					
DvD mit Matching	Keine erhebliche Verbesserung				
IDvD	Beschränkte Verbesserung				
RD-Ansatz	Nicht sinnvoll umsetzbar	Keine erhebliche Verbesserung			
IV-Ansatz	Nicht umsetzbar, da kein geeignetes Instrument verfügbar				
Changes-in-Changes	Verbesserungspotenzial				
Synthetische KG	Verbesserungspotenzial			Verbesserungspotenzial, jedoch evtl. nicht umsetzbar	
Strukturelle/Ex-Ante-Methoden					
Suchmodelle	Beschränkte Verbesserung	Verbesserungspotenzial	Verbesserungspotenzial	Beschränkte Verbesserung	
CGE-Modelle	Prinzipiell geeignet, allerdings langer zeitlicher Vorlauf erforderlich				

Anmerkung: Die Verbesserungspotenziale beziehen sich auf den Vergleich mit dem DvD-Ansatz der ersten Evaluationsrunde.

Die **Implementierung** all dieser Methoden und vor allem Kombinationen derselben ist jedoch mit einem beträchtlichen Zeitaufwand verbunden, der einen Projektrahmen von sechs bis zwölf Monaten in jedem Fall übersteigen dürfte und mindestens mit einem Zeitraum von 18 Monaten zu veranschlagen wäre. Dies gilt umso mehr, da es für einige der Ansätze wie beispielsweise das CiC-Verfahren bislang nur sehr wenige Erfahrungen bei der Umsetzung gibt.

6 Literatur

- Abadie, A. und J. Gardeazabal (2003), The Economic Costs of Conflict: A Case Study of the Basque Country, *American Economic Review* 93, 112-132.
- Abadie, A., Diamond, A. und J. Hainmueller (2010), Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, *Journal of the American Statistical Association* 105, 493-505.
- Abadie, A., Diamond, A. und J. Hainmueller (2012), Comparative Politics and the Synthetic Control Method, Unveröffentlichtes Manuskript. University of Harvard.
- Ahn, T., Arcidiacono, P. und W. Wessels (2011), The Distributional Impacts of Minimum Wage Increases when Both Labor Supply and Labor Demand are Endogenous, *Journal of Business and Economic Statistics* 29, 12-23.
- Albrecht, J. W. und B. Axell (1984), An Equilibrium Model of Search Unemployment, *Journal of Political Economy* 92, 824-840.
- Allegretto, S. A., Dube, A. und M. Reich (2011), Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data, *Industrial Relations* 50, 205-240.
- Allegretto, S. A., Dube, A., Reich, M. und B. Zipperer (2013), Credible Research Designs for Minimum Wage Studies, Working Paper No. 148-13. Berkeley.
- Angrist, J. (1990), Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, *The American Economic Review* 80, 313-336.

- Angrist, J. und A.B. Krueger (2001), Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments, *Journal of Economic Perspectives* 15, 69–85.
- Arntz, M., Boeters, S., Gürtzgen, N. und S. Schubert (2008), Analysing Welfare Reform in a Microsimulation-CGE Model - The Value of Disaggregation, *Economic Modelling* 25, 422-439.
- Athey, S. und G. Imbens (2006), Identification and Inference in Nonlinear Difference-In-Differences Models, *Econometrica* 74, 431-497.
- Bergemann, A., Fitzenberger, B. und S. Speckesser (2009), Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences, *Journal of Applied Econometrics* 24, 797-823.
- Blundell, R. W. und S. Bond (1998), Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics* 87, 115-143.
- Blundell, R. und M. Costa Dias (2009), Alternative Approaches to Evaluation in Empirical Microeconomics, *Journal of Human Resources* 44, 565-640.
- Böhringer, C. und W. Wiegard (2003), Methoden der angewandten Wirtschaftsforschung: eine Einführung in die numerischer Gleichgewichtsanalyse, ZEW-Diskussionspapier 03-03. Mannheim.
- Böhringer, C., Boeters, S. und M. Feil (2005), Taxation and Unemployment: An Applied General Equilibrium Approach for Germany, *Economic Modelling* 22, 81-108.
- Boeters, S. und L. Savard (2011), The Labour Market in CGE-Models, ZEW-Diskussionspapier 11-079. Mannheim.

- Bontemps, C., Robin, J. M. und G. J. van den Berg (1999), An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms, *International Economic Review* 40, 1039–1074.
- Bowlus, A. J., Kiefer, N. M. und G. R. Neumann (1995), Estimation of Equilibrium Wage Distributions with Heterogeneity, *Journal of Applied Econometrics* 10, S119-S131.
- Bowlus, A. J., Kiefer, N. M. und G. R. Neumann (2001), Equilibrium Search Models and the Transition from School to Work, *International Economic Review* 42, 313-343.
- Bunzel, H., Christensen, B. J., P. Jensen, Kiefer, N. M., Korsholm, L., Muus, L., Neumann, G. R. und M. Rosholm (2001), Specification and Estimation of Equilibrium Search Models, *Review of Economic Dynamics* 4, 90-126.
- Burdett, K. und D. T. Mortensen (1998), Wage Differentials, Employer Size, and Unemployment, *International Economic Review* 39, 257-273.
- Cahuc, P. und A. Zylberberg (2004), *Labor Economics*. Cambridge.
- Caliendo, M. und S. Kopeinig (2008), Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys* 22, 31-72.
- Card, D. E. (1992), Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage, *Industrial and Labor Relations Review* 46, 22-37.
- Card, D. und A. B. Krueger (1994), Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, *American Economic Review* 84, 772-93.
- Card, D. E. und A. B. Krueger (1995), *Myth and Measurement: The New Economics of the Minimum Wage*. Princeton.

- Chowdry, H., Meghir, C. und J. Shaw (2009), Investigating the Impact of the Minimum Wage Regime on the Labour Market Behaviour of Young Workers, Institute for Fiscal Studies. London.
- Chung, S., Lee, J. und T. Osang (2013), Did China Tire Safeguard Save U.S. Workers? Unveröffentlichtes Manuskript. University of Memphis.
- Dehejia, R. H. und S. Wahba (1999), Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, *Journal of the American Statistical Association* 94, 1053-1062.
- Diamond, P. (1971), A Model of Price Adjustment, *Journal of Economic Theory* 3, 156-168.
- Dickens, R., Machin, S. und A. Manning (1999), The Effects of Minimum Wages on Employment: Theory and Evidence from Britain, *Journal of Labor Economics* 17, 1–22.
- Dickens, R., Riley, R. und D. Wilkinson (2011), The UK Minimum Wage at Age 22: A Regression Discontinuity Approach, Working Paper 21-2011. University of Sussex.
- DiNardo, J. und D. Lee (2004), Economic Impacts of New Unionization on Private Sector Employers: 1984–2001, *Quarterly Journal of Economics* 119, 1383-1441.
- Dolton, P., Rosazza Bondibene, C. R. und J. Wadsworth (2010). The UK National Minimum Wage in Retrospect, *Fiscal Studies* 31, 509-534.
- Dolton, P., Rosazza Bondibene, C. R. und J. Wadsworth (2012a), Employment, Inequality and the UK National Minimum Wage over the Medium-Term, *Oxford Bulletin of Economics and Statistics* 74, 78-106.
- Dolton, P., Rosazza Bondibene, C. R. und M. Stops (2012b), The Spatial Analysis of the Employment Effect of the Minimum Wage in a Recession - the Case of the UK 1999-2010, Research Report Commissioned by the Low Pay Commission. London.

- Dube, A., Lester, T. W. und M. Reich (2010) Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties, *Review of Economics and Statistics* 92, 945-964.
- Dube, A. und B. Zipperer (2013), Pooled Synthetic Control Estimates for Recurring Treatments: An Application to Minimum Wage Case Studies. Unveröffentlichtes Manuskript. University of Massachusetts. Amherst.
- Eckstein, Z. und K. I. Wolpin (1990), Estimating a Market Equilibrium Search Model from Panel Data on Individuals, *Econometrica* 58, 783-808.
- Eckstein, Z. und G.J. van den Berg (2007), Empirical Labor Search – A Survey, *Journal of Econometrics* 136, 531-564.
- Elhorst, J. P. (2010), Spatial Panel Data Models, in: Handbook of Applied Spatial Analysis, 377-407. Berlin, Heidelberg.
- Fitzenberger, B. und A. Garloff (2007), Labor Market Transitions and the Wage Structure in Germany, *Jahrbücher für Nationalökonomie und Statistik* 227, 115-152.
- Fitzenberger, B. und A. Garloff (2008), Unemployment, Labor Market Transitions, and Residual Wage Dispersion, *Scottish Journal of Political Economy* 55, 561-590.
- Fitzenberger, B. (2009), Anmerkungen zur Mindestlohndebatte: Elastizitäten, Strukturparameter und Topfschlagen, *Zeitschrift für Arbeitsmarktforschung* 42, 85-92.
- Flinn, C. J. (2006), Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates, *Econometrica* 74, 1013-1062.
- Flinn, C. J. und J. Mabli (2008), On-the-Job Search, Minimum Wages, and Labor Market Outcomes in an Equilibrium Bargaining Framework, Arbeitspapier Nr. 91, Collegio Carlo Alberto, Turin.
- Franz, W. (2013), Arbeitsmarktökonomik. Heidelberg.

- Frings, H., vom Berge, P. und A. R. Paloyo (2013), Minimum Wages and Employment: A Spatial Identification Strategy for Germany, Beitrag zur Jahrestagung des Vereins für Socialpolitik 2013.
- Garloff, A. A. (2010), Minimum Wages, Wage Dispersion and Unemployment in Search Models – A Review, *Zeitschrift für Arbeitsmarktforschung – Journal for Labour Market Research* 43, 145-167.
- Gorry, A. (2013), Minimum Wages and Youth Unemployment, *European Economic Review* 64, 57-75.
- Holzner, C. und A. Launov (2010), Search Equilibrium and Social and Private Returns to Education, *European Economic Review* 54, 39-59.
- Green, D. A. und K. Harrison (2009), Minimum Wage Setting and Standards of Fairness, IFS Working Paper W10/09.
- Gürtzgen, N. (2003), Makroevaluation mit berechenbaren allgemeinen Gleichgewichtsmodellen, in: Hagen, T. und A. Spermann, Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung. ZEW-Wirtschaftsanalysen Bd. 74. Baden-Baden.
- Hahn, J., Todd, P. E. und W. van der Klaauw (2001), Identification and Estimation of Treatment Effects with a Regression Discontinuity Design, *Econometrica* 69, 201-209.
- Hamermesh, D. S. (1993), Labor Demand. Princeton.
- Heckman, J. J., Ichimura, H. und P. E. Todd (1997), Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *The Review of Economic Studies* 64, 605-654.
- Heckman, J. J. und S. Navarro-Lozano (2004), Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models, *The Review of Economics and Statistics* 86, 30-57.
- IAB/RWI/ISG (2011), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Bauhauptgewerbe. Nürnberg, Essen, Köln.

- IAQ (2011a), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Gebäudereinigung. Duisburg.
- IAQ (2011b), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Wäschereidienstleistungen im Objektkundengeschäft. Duisburg.
- IAW (2011a), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Maler und Lackiererhandwerk. Tübingen.
- IAW (2011b), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Elektrohandwerk. Tübingen.
- IAW (2011c), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Pflege. Tübingen.
- Imbens, G. und J. Wooldridge (2009), Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature* 47, 5-86.
- Kim, T. und L. J. Taylor (1995), The Employment Effect in Retail Trade of California's 1988 Minimum Wage Increase, *Journal of Business and Economic Statistics* 13, 175-182.
- Klepper, G., Lorz, J. O., Stähler, F., Thiele, R. und M. Wiebelt (1994), Empirische allgemeine Gleichgewichtsmodelle – Struktur und Anwendungsmöglichkeiten, *Jahrbücher für Nationalökonomie und Statistik* 213, 513-544.
- Kluve, J. (2010), The Effectiveness of European Active Labor Market Programs, *Labour Economics* 17, 904-918.
- Krueger, A. B. (1999), Experimental Estimates of Education Production Functions, *Quarterly Journal of Economics* 114, 497–532.
- Launov, A., Wolff, J. und S. Klasen (2004), Unemployment Benefits and Unemployment Rates of Low-Skilled and Elder Workers in West Germany: A Search Equilibrium Approach, IZA Diskussionspapier 1161, Bonn.

- Lechner, M. (2010), The Estimation of Causal Effects by Difference-in-Difference Methods, *Journal of Foundations and Trends in Econometrics* 4, 165–224
- Lechner, M. (2011), The Estimation of Causal Effects by Difference-in-Difference-Methods, Diskussionspapier 2010-28. St. Gallen.
- Lee, D. S. und T. Lemieux (2010), Regression discontinuity designs in economics, *Journal of Economic Literature* 48, 281–355.
- Lemos (2005), Political Variables as Instruments for the Minimum Wage, *Contributions to Economic Analysis & Policy* 4, Article 16.
- Machin, S. und A. Manning (1994), The Effects of Minimum Wages on Wage Dispersion and Employment: Evidence from the UK Wages Councils, *Industrial and Labor Relations Review* 47, 319-329.
- Magruder, J. R. (2013), Can Minimum Wages Cause a Big Push? Evidence from Indonesia, *Journal of Development Economics* 100, 48-62.
- Melly, B. und G. Santangelo (2013), The Changes-in-Changes Model with Covariates, Vortrag auf der Statistischen Woche 2013 an der Freien Universität Berlin.
- Möller, J. (2012), Minimum Wages in German Industries – What Does the Evidence Tell Us so Far?, *Zeitschrift für Arbeitsmarktforschung – Journal for Labour Market Research* 45, 187-199.
- Moore, C. T. (2008), Estimating Policy Effects Using Spatial Regression Discontinuity: The Case of New Jersey's Minimum Wage Increase. Poster presented at the Annual Department of Educational Psychology Graduate Student Research Day. Minneapolis.
- Neumark, D. und W. Wascher (1992), Evidence on Employment Effects of Minimum Wages and Subminimum Wage Provisions: Panel Data on State Minimum Wage Laws, *Industrial and Labor Relations Review* 46, 55–81.

- Neumark, D. und W. Wascher (2006), Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research, NBER Working Paper 12663. Cambridge.
- Neumark, D., Salas, J. M. I. und W. Wascher (2013), Revisiting the Minimum Wage-Employment Debate: Throwing out the Baby with the Bathwater?, IZA Diskussionspapier 7166. Bonn.
- Nguyen, C. V. (2013), Do Minimum Wage Increases Matter to Firm Profitability? The Case of Vietnam, erscheint in: *Journal of International Development*.
- Olsen, A. (2011), The Short Run Effects of Age Based Youth Minimum Wages in Australia: A Regression Discontinuity Approach, Working Paper. Melbourne.
- Rosenbaum, P. R. und R. Rubin (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70, 41-55.
- Sabia, J., Burkhauser, R. und B. Hansen (2012), Are Minimum Wage Effects Always Small? Evidence from a Case Study of New York State, *Industrial and Labor Relations Review* 65, 433-444.
- Sen, A., Rybczynski, K. und C. van de Waal (2011), Teen Employment, Poverty, and the Minimum Wage: Evidence from Canada, *Labour Economics* 18, 36-47.
- Sorkin, I. (2013), Are There Long-Run Effects of the Minimum Wage?, Unveröffentlichtes Manuskript, University of Michigan.
- Statistisches Bundesamt (2008), Klassifikation der Wirtschaftszweige – mit Erläuterungen, Wiesbaden.
- Tatsiramos, K. und J.C. van Ours (2012), Labor Market Effects of Unemployment Insurance Design, IZA Diskussionspapier Nr. 6950. Bonn.

- Stewart, M. B. (2002), Estimating the Impact of the Minimum Wage Using Geographical Wage Variation, *Oxford Bulletin of Economics and Statistics* 64 (Supplement), 583-605.
- Van den Berg, G. J. (1999), Empirical Inference with Equilibrium Search Models of the Labour Market, *The Economic Journal* 109, F283–F306.
- Van den Berg, G. J. und G. Ridder (1998), An Equilibrium Search Model of the Labor Market, *Econometrica* 66, 1183–1221.
- Van den Berg, G. J. und A. Uhlendorff (2013), Economic Job Search and Decision-Making Models, Unveröffentlichtes Manuskript. Universität Mannheim.
- Van Vuuren, A., van den Berg, G. J. und G. Ridder (2000), Measuring the Equilibrium Effects of Unemployment Benefits Dispersion, *Journal of Applied Econometrics* 15, 547-574.
- Vella, F. und M. Verbeek (1999), Estimating and Interpreting Models with Endogenous Treatment Effects, *Journal of Business and Economic Statistics* 17, 473–478.
- Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*. 2. Auflage. Cambridge.
- ZEW (2011a), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Dachdeckerhandwerk. Mannheim.
- ZEW (2011b), Abschlussbericht zur Evaluation bestehender gesetzlicher Mindestlohnregelungen: Branche Abfallwirtschaft. Mannheim.

7 Anhang

Background information on the evaluation of minimum wages in Germany

Minimum wages in Germany: In contrast to many other industrialized countries, Germany has not (yet) introduced a comprehensive national statutory minimum wage. Instead, (nationwide) minimum wages have been implemented in selected industries only. In a number of cases, the minimum wages vary by worker groups and between East and West Germany. The institutional prerequisites for such an introduction are that a specific industry is subject to the German "Posting of Workers Act"³¹ and that a minimum wage collective bargaining agreement has been negotiated between the industry's respective collective bargaining parties. Once these conditions are met, the Federal Ministry of Labor may decide to declare the agreement to be generally binding for the whole industry.

First Round of Minimum Wage Evaluations: In 2010, the Federal Ministry of Labor initiated a scientific evaluation of the introduction of industry-specific minimum wages in eight different industries (Construction, Roofing, Painters and Varnishers, Electrical Trade, Commercial Cleaning, Industrial Laundries, Waste Management and Nursing Care). The overall aim of this evaluation was to assess the impact of the introduction of minimum wages on employment, wages and the degree of competition in the respective industries. The evaluations have been conducted for each industry separately resulting in eight studies. In 2014, the Ministry plans to launch a second round of evaluations that shall make use of more refined or alternative evaluation methods as well as of

³¹ The "Posting of Workers Act" forces national as well as foreign employers (temporarily posting workers to German employers) to comply with national minimum wage and labor legislations. When the law came into effect in 1996, its implementation was confined to construction and several sub-construction industries, with the primary aim to protect these industries from foreign low-wage competition.

more comprehensive data. It will also include those industries where a minimum wage was only introduced in 2011 or afterwards.

Methodological Approach: Despite the heterogeneity of the industries investigated, difference-in-difference estimates were the dominating approach in the first round of evaluations. Identification relied on two assumptions: (1) treatment and control group experience common trends and (2) the absence of spillover effects between both groups, which may occur due to substitution or scale effects in labor demand or due to direct spillover effects in wage setting. While the studies provided some first quasi-experimental evidence on the effects of minimum wages in Germany, the empirical strategy outlined above is open to several criticisms:

Potential concerns and prospects for further improvements:

Choice of control groups: Depending on the data used, the control groups that were selected can be broadly categorized into (1) workers/employers from the same industry for whom the minimum wage is not binding and (2) workers/employers from a different industry. Throughout the studies it has been argued that the former are more likely to fulfill the common trends assumption, whereas the latter are more likely to meet the absence of spillover effects. For those studies that made use of both categories, a particular concern is that the results appear to be fairly sensitive to the selected control group. Thus far, no attempt has been made to optimally combine information on potentially available, different control groups. Further, given the likely relevance of spillover effects, there has been no systematic attempt to validate either control group category by quantifying possible substitution or scale effects.

Policy endogeneity and anticipation effects: Given the institutional requirements spelled out above, the introduction of a minimum wage strongly hinges on the bargaining parties' initiative. This may imply that unions and employer associations of those industries that are particularly prone to unfavorable developments are the most likely to successfully lobby for a minimum wage.

Moreover, a further concern is the issue of potential anticipation effects, as the bargaining parties may strategically alter their behavior prior to the introduction of a minimum wage.

No explicit modeling of the labor market structure: The fact that the empirical strategy adopted is a reduced-form approach implies that the studies remain silent about the underlying transmission mechanisms of minimum wages. Given that negative employment effects may be reconciled with both a competitive and a monopsonistic labor market, the analyses do not allow any conclusions about the underlying labor market structure.

General equilibrium effects: Due to its industry-specific focus, the evaluation approach was only partial-equilibrium in nature, which implies that spillover effects between different industries and macro feed-back effects, such as repercussions on the price level, the government budget etc. have been typically neglected.