

„Outputorientierte Evaluierung öffentlich geförderter FTI-Programme – Möglichkeiten und Grenzen“

Studie im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie

Christian Reiner (Joanneum Research, JR)

Stefanie Smoliner (Zentrum für Soziale Innovation, ZSI)

MÄRZ 2012

„Outputorientierte Evaluierung öffentlich geförderter FTI-Programme – Möglichkeiten und Grenzen“

Studie im Auftrag des Bundesministeriums für Verkehr, Innovation und Technologie

Projektleitung:

Karl Blecha (GFF)

Günter Hillebrand (GFF)

Wolfgang Polt (Joanneum Research)

Projektteam:

Wolfgang Polt (Joanneum Research)

Christian Reiner (Joanneum Research)

Klaus Schuch (ZSI)

Stefanie Smoliner (ZSI)



INHALTSVERZEICHNIS

EXECUTIVE SUMMARY	7
1 EINLEITUNG	11
2 INNOVATIONSPOLITISCHER KONTEXT VON FTI-EVALUIERUNGEN IN ÖSTERREICH.....	17
2.1. Institutionen, Akteure und Strategien.....	17
2.2. F&E-Förderungen im Überblick	19
2.3. FFG-Förderungen und thematische Orientierung des BMVIT im Überblick.....	27
3 FTI-EVALUIERUNG IN ÖSTERREICH: STATUS QUO UND INSTITUTIONELLE HERAUSFORDERUNGEN	31
3.1. Bestands- und Bedürfnisanalyse österreichischer FTI-Evaluierungen	31
3.1.1 Viele und auch gute Evaluierungen.....	31
3.1.2 Unbefriedigte Bedürfnisse und Verbesserungspotenzial.....	32
3.2. Institutionelle Herausforderungen, Evaluierungsdesign und Zielkonflikte	36
3.2.1 Themenmanagement	36
3.2.2 Wirkungsorientierte Haushaltsführung	42
4 NATIONALE UND INTERNATIONALE EVALUIERUNGSTRENDS VON INNOVATIONSPOLITISCHEN MAßNAHMEN IM ÜBERBLICK	51
4.1. Ausgewählte österreichischen Evaluierungen von 2005-2011	51
4.2. Ergebnisse aus der Analyse der Programmdokumente und Evaluationen	55
4.2.1 Ziele	56
4.2.2 Datenlage	58
4.2.3 Evaluierungsmethoden.....	59
4.3. Evaluation innovationspolitischer Maßnahmen im internationalen Vergleich.....	63
4.3.1 Österreichische FTI-Evaluation im internationalen Vergleich: Die Ergebnisse des INNO-Appraisal-Projekts.....	64
4.3.2 Kosten-Nutzen Analyse als angelsächsisches Spezifikum	67
4.3.3 Überblick zu aktuellen ökonometrischen Evaluierungen innovations- politischer Maßnahmen	73
5 EVALUIERUNGSMETHODEN IM FTI-KONTEXT: THEORIE, INTERNATIONALE FALLBEISPIELE UND ANWENDUNG IN ÖSTERREICH.....	79
5.1. Qualitative Methoden in FTI-Wirkungsanalysen	82
5.1.1 Gemeinsamkeiten qualitativer Methoden	84
5.1.2 Die qualitative soziale Netzwerkanalyse.....	84
5.1.3 Aktuelle Trends in der Fokusgruppenforschung.....	88
5.2. Quantitative Methoden	94
5.2.1 Entwicklung der mikroökonomischen Programmevaluierung.....	95

5.2.2	Rubin Causal Model und das Evaluationsproblem	98
5.2.3	Fundamentale Identifikationsstrategien und Sample Selection Bias	104
5.2.4	Ausgewählte mikroökonomische Methoden zur Programmevaluierung	124
5.2.5	Grenzen ökonomischer Modellierung	136
5.3.	Bestimmungsgründe der Methodenwahl	137
6	SOZIO-ÖKONOMISCHE WIRKUNGSDIMENSIONEN VON FTI-MAßNAHMEN	141
6.1.	Definition sozio-ökonomischer Wirkungsdimensionen.....	141
6.2.	Welche sozio-ökonomischen Wirkungsdimensionen werden bei der Evaluierung von FTI-Massnahmen generell erhoben?	142
6.3.	Herausforderungen durch den Einbezug nicht-ökonomischer Wirkungsdimensionen bei der Evaluierung von FTI-Massnahmen.....	143
6.4.	Wirkungsdimensionen messbar machen – jenseits der ökonomischen Dimension	145
6.5.	Messung nicht-ökonomischer Wirkungsdimensionen	149
7	DATENVERFÜGBARKEIT ALS KRITISCHER ENGPASSFAKTOR BEI EVALUIERUNGEN.....	151
7.1.	Die Komplementarität von Daten und Methoden – eine produktions-theoretische Perspektive	151
7.2.	Die Datenlage in Österreich.....	153
7.3.	Es geht auch anders – internationale Vergleichsbeispiele	157
7.4.	Plädoyers für die Verwendung amtlicher und nicht-amtlicher Sekundärstatistiken	162
8	RESÜMEE UND EMPFEHLUNGEN	167
	REFERENZEN	173
	ANHANG A : INDIKATOREN FÜR VERSCHIEDENE SOZIO-ÖKONOMISCHE WIRKUNGSDIMENSIONEN	179
	ANHANG B: INTERVIEWPARTNERINNEN	183
	ANHANG C: INTERVIEWLEITFADEN	185

Abbildungsverzeichnis

Abbildung 1: Anteil des Staates an den volkswirtschaftlichen F&E-Ausgaben (GERD) in % des BIP (2008)	21
Abbildung 2: Prozentualer Anteil der vom Staat finanzierten F&E-Ausgaben des Unternehmenssektors	22
Abbildung 3: Entwicklung von BIP und F&E-Ausgaben 2006-2011 (2006=100)	23
Abbildung 4: Direkte und indirekte Subvention von unternehmerischer F&E (BERD) 2008 (oder letztes verfügbares Jahr) in % des BIP	24
Abbildung 5: Anteil innovativer Unternehmen mit Innovationsförderung in % aller technologisch innovativen Unternehmen	25
Abbildung 6: Innovative Unternehmen in Österreich mit Innovationsförderung differenziert nach Größenklasse und Fördergeber	26
Abbildung 7: Qualität der Evaluationen	35
Abbildung 8: Klassische Programmevaluierung	41
Abbildung 9: Steigende Komplexität durch Portfolio- bzw. Themenevaluierung	41
Abbildung 10: Budgetstruktur und WHF mit zugeordneten Wirkungszielen für den FTI-Sektor nach Entwürfen des BMVIT	45
Abbildung 11: Tag Cloud der sozioökonomischen Wirkungsdimensionen ausgewählter FTI-Programme	61
Abbildung 12: Angewandte Methoden bei FTI-Evaluierungen in Österreich und Vergleichsländern	65
Abbildung 13: Inhaltliche Ausrichtung der Evaluationen	67
Abbildung 14: Nettonutzen der Forschungsergebnisse	71
Abbildung 15: Kosten- und Nutzenkomponenten von Projekt SAR 01/05	72
Abbildung 16: Determinanten des F&E-Investitionsverhalten ab und dessen Auswirkungen?	80
Abbildung 17: Das Eisbergmodell: Nur ein Teil der Fördereffekte lässt sich "einfach" beobachten	81
Abbildung 18: FTI-Evaluationsmethoden im Überblick	82
Abbildung 19: Potentielle Einflussfaktoren bei der Durchführung von Fokusgruppen	93
Abbildung 20: Entwicklung der durchschnittlichen Noten von StudentInnen	101
Abbildung 21: Der Querschnittsvergleich und mögliche Kontrollgruppen	107
Abbildung 22: Idealtypischer Selektionsprozess bei F&E-Programmen	108
Abbildung 23: Vorher-Nachher-Schätzer bei positivem kontrafaktischem Trend	116
Abbildung 24: Vorher-Nachher-Schätzer bei negativem kontrafaktischem Trend	117
Abbildung 25: Der Differenz-von-Differenzen-Ansatz	118
Abbildung 26: Die kontrafaktische Frage im FFG-Wirkungsmonitoring 2010	119
Abbildung 27: Fragen im FFG-Wirkungsmonitoring zur Ermittlung des Fördermultiplikators	121
Abbildung 28: Direkte Frage nach der kontrafaktischen Situation als Identifikationsannahme	122
Abbildung 29: Klassifikation mikroökonomischer Methoden nach der Notwendigkeit zur theoriegeleiteten Modellierung	125
Abbildung 30: Common Support Assumption ist erfüllt	131

Abbildung 31: Common Support Assumption ist möglicherweise verletzt.....	131
Abbildung 32: Korrektur beim Kontrollfunktionsansatz	135
Abbildung 33: Vor- und Nachteile von Matching- und Kontrollfunktionsansätzen im Vergleich ...	136
Abbildung 34: Evaluationsmethoden und Analyseebene	139
Abbildung 35: Der Zusammenhang von Methode und Anzahl der Programmteilnehmer	139
Abbildung 36: Arbeitsschritte einer Evaluierung.....	146
Abbildung 37: Evaluatorische Information wird mittels einer limitationalen Produktionsfunktion generiert.....	152
Abbildung 38: Datensätze zur Evaluierung FTI-politischer Maßnahmen	156
Abbildung 39: Datenstruktur bei Czarnitzki et al. (2003).....	158
Abbildung 40: Datenstruktur bei Arvanitis et al. (2010)	159
Abbildung 41: Datenstruktur bei Aerts und Czarnitzki (2006).....	160
Abbildung 42: Datenstruktur bei Marino und Parrotta (2010).....	160
Abbildung 43: Datenstruktur bei Einiö (2009).....	161

Tabellenverzeichnis

Tabelle 1: Niveau und Trends staatlicher Subventionen im internationalen Vergleich (in % des BIP)	20
Tabelle 2: Absolute Häufigkeiten geförderter und nicht geförderter innovativer Unternehmen in Österreich (2008)	26
Tabelle 3: Der "Delivery Gap"	33
Tabelle 4: Der "Customer Gap"	33
Tabelle 5: Der "Management Gap"	34
Tabelle 6: Instrumentenkoffer Entwurf 1: Förderinstrumente und Förderkonditionen (Stand August 2011)	38
Tabelle 7: Instrumentenkoffer Entwurf 2: Instrumente und zugeordnete Programme (Stand August 2011)	39
Tabelle 8: Entwurf zu Zielen, Maßnahmen und Indikatoren zur WHF im BMVIT	47
Tabelle 9: Österreichische FTI-Evaluierungen 2005-2011 im Überblick.....	52
Tabelle 10. Übersicht der analysierten Programmevaluationen	55
Tabelle 11: Kosten-Nutzen-Analysen von NIST-Projekten	69
Tabelle 12 Auswahl rezenter ökonomischer, internationaler Evaluationen von innovationspolitischen Interventionen.....	75
Tabelle 13: Quantitative und qualitative Verfahren der Netzwerkforschung im Vergleich	86
Tabelle 14: Beobachtbare und unbeobachtbare Situationen.....	99
Tabelle 15: F&E-Intensität bei geförderten und nicht geförderten Unternehmen im deskriptiven Vergleich.....	111
Tabelle 16: Antworten auf die kontrafaktische Frage 2006.....	120
Tabelle 17: Einteilung mikroökonomischer Evaluationsverfahren	126

Tabelle 18: Ergebnisse der FlexCIM-Evaluierung	132
Tabelle 19: Ergebnisse des Matchingverfahrens der CTI-Evaluierung (Nearest Neighbour Matching)	133
Tabelle 20: Innovationspolitische Instrumente und adäquate Innovationsmethoden	138
Tabelle 21: Die Bandbreite sozio-ökonomischer Wirkungsdimensionen.....	142
Tabelle 22: Öffentliche Finanzierung der Ausgaben für F&E im Unternehmenssektor 2009 (Firmeneigener Bereich).....	154

Boxenverzeichnis

Box 1: Kommentierter Methodenüberblicke zur Evaluation von FTI-Maßnahmen	12
Box 2: Aktuelle Stellungnahmen zur Evaluierung von FTI-Maßnahmen	19
Box 3: Fallbeispiel Government Performance and Results Act (GPRA), USA, 1993	43
Box 4: Ziele der zweiten Etappe der Haushaltsrechtsreform	44
Box 5: Beispiel für eine Einteilung FTI-politischer Zielsetzungen auf verschiedene Zielebenen	57
Box 6: Internationale Unterschiede bei der Evaluation von FTI-Politik.....	64
Box 7: ATP und NIST als Beispiele für innovationspolitische Institutionen, die KNA als Evaluationsmethode anwenden	68
Box 8: Fallbeispiel KNA in Australien	70
Box 9: Kennzeichen qualitativer Evaluationsforschung nach Mayring (2002)	83
Box 10: Wichtige Beiträge qualitativer Verfahren zur Netzwerkforschung	86
Box 11: Beispiel für ein quantitativ-qualitatives Fokusgruppensdesign	92
Box 12: Erfassung heterogener Maßnahmeneffekte. Fallbeispiel Universitätspolitik.....	101
Box 13: Rückwirkung industriepolitischer Maßnahmen auf andere Unternehmen – ein stilisiertes Beispiel	104
Box 14: Sample Selection Probleme bei der Evaluation spanischer FTI-Subventionen.....	111
Box 15: Ökonometrische Evaluation der FFG-Förderungen anhand eines Vorher-Nachher-Schätzers	114
Box 16: Die Identifikationsannahmen des FFG-Wirkungsmonitorings.....	119
Box 17: Einwände gegen die direkte Frage als Identifikationsstrategie.....	124
Box 18: Matchingansätze in der FTI-Evaluierung	132
Box 19: Gründe für die Dominanz der ökonomischen Wirkungsdimensionen.....	142
Box 20: Beispiele für Indikatoren zur Bestimmung wissenschaftlicher Qualität auf der Individualebene	148
Box 21: Fallbeispiel Deutschland.....	158
Box 22: Fallbeispiel Schweiz	159
Box 23: Fallbeispiel Flandern.....	160
Box 24: Fallbeispiel Dänemark	160
Box 25: Fallbeispiel Finnland	161
Box 26: Verbesserung der Datenlage: Dateninitiativen des RFTE und des Rechnungshofs	165

Executive Summary

Konsolidierungsbedarf bei den öffentlichen Budgets, berechtigte Forderungen nach möglichst effizienter Verwendung öffentlicher Mittel und das Streben der Politik nach möglichst großer Wirkung (Outputs) durch die eingesetzten Fördermittel machen es notwendig, die Messung von Output und Effizienz von FTI-Förderungen immer weiter zu verbessern.

Die Bedeutung möglichst fundierten Wissens über die Effekte von FTI-Förderungen ist dabei gerade in Österreich, das einen der höchsten Anteile staatlicher Förderung unternehmensbezogener F&E in Europa aufweist, hoch und wird bereits seit einigen Jahren intensiv diskutiert – etwa in den Diskussionen um die ‚wirkungsorientierte Haushaltsführung‘ oder um das Portfolio-Management von FTI-Programmen.

Vor diesem Hintergrund hat das BMVIT bei der Gesellschaft zur Förderung der Forschung (GFF) die vorliegende Studie „Outputorientierte Evaluierung öffentlich geförderter FTI-Programme: Möglichkeiten und Grenzen“ in Auftrag gegeben, die – von JOANNEUM RESEARCH (JR) und dem Zentrum für Soziale Innovation (ZSI) durchgeführt – die Politik unterstützen soll, um die Wirkungen ihrer Förderungen besser abschätzen zu können.

In dieser Studie wird das österreichische Evaluierungssystem im FTI-Bereich mit Schwerpunkt auf FFG- und BMVIT-Programme vor dem Hintergrund internationaler Vergleiche und nationaler Herausforderungen analysiert. Neben institutionellen Aspekten der Evaluierung werden dabei insbesondere die Evaluierungsmethoden und die Möglichkeiten und Notwendigkeiten ihrer Weiterentwicklung dargestellt.

Die wichtigsten Befunde der Studie zur aktuellen österreichischen Evaluierungspraxis sind folgende:

- Evaluierungen im FTI-Bereich finden in Österreich im internationalen Vergleich häufig statt und sind im Allgemeinen von guter Qualität.
- Defizite bestehen vor allem (i) bei ex-post Evaluierungen, Portfolioevaluierungen und Effizienzanalysen, die vergleichsweise selten und meist vor dem Hintergrund beschränkter Datenverfügbarkeit durchgeführt werden sowie (ii) in der Verbindlichkeit der Umsetzung von Evaluierungsergebnissen.
- Für eine bessere Effizienz- und Effektivitätsmessung wird sowohl im Bereich der qualitativen aber vor allem auch in der Anwendung quantitativer Methoden ein deutlicher Entwicklungsbedarf gesehen. Dieser liegt bei den qualitativen Methoden etwa in neuen Fokusgruppen-Designs und qualitativer sozialer Netzwerkanalyse, bei den quantitativen Methoden vor allem in komplexeren mikroökonomischen Verfahren wie etwa dem Matching-Ansatz und Kontrollgruppenverfahren.

- Für den Einsatz der meisten dieser Verfahren ist die Verfügbarkeit von Daten auf der Mikroebene (Projekt, Unternehmen) Voraussetzung. Diese ist allerdings zurzeit in Österreich noch nicht in ausreichendem Maß gegeben. Ein internationaler Vergleich macht deutlich, dass die Situation in vergleichbaren OECD-Staaten, wie z.B. Deutschland, Schweiz oder Finnland deutlich besser ist. Hier hat Österreich eindeutig Nachholbedarf.
- Die Studie kommt zu einem differenzierten Urteil was die Möglichkeiten und Grenzen der o.a. Methoden zur Wirkungsmessung angeht: Während man mit den beschriebenen Methoden deutliche Fortschritte in der Messung von ‚Additionalität‘ (‚Wäre der Effekt ohne die Maßnahme nicht/schwächer/anders aufgetreten?‘) machen könnte, stoßen manche Hoffnungen auf präzise Identifikation und Zurechnung von Outputs (etwa Wachstums- und Beschäftigungseffekte i.w.S.) auf prinzipielle methodische Grenzen. Solche weitreichenden evaluativen Fragestellungen sollten daher von der Politik nur unter Bedachtnahme auf das theoretisch mögliche gestellt werden.

Zudem ist zu beachten, dass die zunehmende Forderung nach (schon ex-ante identifizierbaren) Outputs in Widerspruch zur Forderung nach vermehrter Förderung risikoreicher F&E Projekte geraten kann.

Zusammenfassend ergeben sich folgende Handlungsempfehlungen:

Weiterentwicklung der Evaluationsmethoden zur besseren Wirkungsmessung. Für die österreichische Situation bedeutet dies den vermehrten Einsatz von avancierteren quantitativen Verfahren, die stärker als die bisher angewandten auf Wirkungsmessung abzielen, sowie von neueren qualitativen Methoden. Hier ist ein ausgewogener Methodenmix von qualitativen und quantitativen Methoden anzustreben.

EvaluatorInnen sollten durch die Ausschreibungsbedingungen und die evaluativen Fragestellungen ermutigt werden, innovative neue methodische Ansätze auszuprobieren und zur kontinuierlichen Weiterentwicklung bestehender Methoden beizutragen.

- **Verbesserung der Datenlage für Evaluierungszwecke.** Die Verbesserung der Datenlage stellt eine wesentliche Vorbedingung für wirkungsorientierte Evaluierung von FTI-Politik dar. Die Situation bezüglich Datenlage ist nicht nur in anderen Ländern (z.B. DEU, FIN, CH), sondern auch in anderen Politikfeldern (Arbeitsmarkt-, Exportstatistiken) in Österreich deutlich besser. Daten, die mit öffentlichen Mitteln erhoben wurden sollten für die Kontrolle der staatlichen Mittelverwendung eingesetzt werden können (wie der Community Innovation Survey oder die F&E-Erhebung) und in Form anonymisierter Mikrodaten zugänglich gemacht werden. Die Möglichkeit zur Verknüpfung mit weiteren Sekundärdatenquellen ist ebenfalls eine notwendige Voraussetzung für die Anwendung von state-of-the-art Methoden der Evaluationsforschung. Gegebenenfalls sind Änderungen der rechtlichen Grundlagen des Datenzugangs in Österreich anzustreben.

- **Verbesserte Abstimmung der Programme und Instrumente durch vermehrte Portfolioevaluierung.** Das österreichische Innovationsförderungssystem ist durch eine Vielzahl an Akteuren und Programmen gekennzeichnet. Doppelförderungen und Förderkonkurrenz sind mögliche unerwünschte Nebenwirkungen eines solchen Fördersystems. Umso wichtiger wird aber die Berücksichtigung der Einbettung des zu evaluierenden Programms in die bestehende Förderlandschaft, d.h. eine Portfolioevaluierung. Diese sollte vermehrt Berücksichtigung finden, um den Instrumentenmix zumindest in einem iterativen Prozess aufeinander abzustimmen.
- **Verbesserung der Qualität und des Nutzens von Evaluierungen durch konzise und akkordierte Formulierung der „Terms of References“ (ToR).** Die in den Ausschreibungen von Evaluierungen formulierten Anforderungen und evaluatorischen Fragestellungen (Terms of References) haben einen zentralen Einfluss auf die Qualität von Evaluierungen. Die Studie kommt zu dem Ergebnis, dass die ToR noch weiter verbessert werden können. Drei Maßnahmen hierzu sind: (1) die Einbeziehung der Förderagenturen in den Prozess der Formulierung der ToR, (2) die klare Arbeitsteilung zwischen Festlegung der ToR durch die Auftraggeber von Evaluierung und die Entscheidung über die Methode durch die EvaluatorInnen sowie (3) realistische Ansprüche an den Umfang der zu beantwortenden Fragen.
- **Klare Zielformulierung in der Programmkonzeption als Voraussetzung für Wirkungsabschätzung.** Wirkungsanalysen brauchen eindeutig definierte und messbare Ziele, die den EvaluatorInnen Aussagen in Bezug auf relevante Veränderungen und Wirkungen ermöglichen. Werden Ziele zu abstrakt formuliert (z.B. Steigerung der Wettbewerbsfähigkeit) erschwert bzw. verunmöglicht dies die Wirkungsabschätzung.
- **Verstärkte Reflexion von sozioökonomischen Indikatoren.** Um die Auswirkungen von Programmen umfassend zu analysieren, sollten auch Wirkungstypen die über ökonomische Effekte hinausgehen (z.B. ökologische Nachhaltigkeit, Nutzerfreundlichkeit, Gender Equality, Public Awareness) in die Wirkungsanalyse einbezogen werden. Hierfür gibt es jedoch noch wenige bis keine allgemein akzeptierten Indikatoren – diese müssen sowohl in Bezug auf die unmittelbaren Ziele sowie den Inhalt und Kontext des Programms formuliert und reflektiert werden. Sollten solche Ziele in Programmen vorkommen bzw. der Evaluierung gewünscht sein, müssten sie dafür im Kontext des jeweiligen Programmes geeignet operationalisiert werden. In der Entwicklung neuer Metriken für die sozio-ökonomischen Indikatoren liegt ebenfalls eine Herausforderung für die Zukunft.
- **Mehr lernen von Evaluierungen durch breite und öffentliche Diskussion.** Um den Einfluss von Evaluierungen zu erhöhen, ist eine breite und möglichst zeitnahe Veröffentlichung von Evaluierungen wünschenswert.

1 Einleitung

„An emphasis on objective publicly accessible evaluations is a distinctive feature of the modern welfare state, especially in an era of limited funds and public demands for accountability“

Heckman et al. (1999:1876)

Die Studie „Outputorientierte Evaluierung öffentlich geförderter FTI-Programme: Möglichkeiten und Grenzen“ analysiert das gegenwärtige österreichische Evaluierungssystem Österreichs im FTI-Bereich mit Schwerpunkt auf FFG und BMVIT-Programme vor dem Hintergrund internationaler Vergleiche und nationaler Herausforderungen. Neben institutionellen Aspekten der Evaluierung werden dabei insbesondere auch die methodischen Aspekte in den Blick genommen.

Diese inhaltliche Orientierung kann aus wirtschaftspolitischer, institutioneller und methodischer Perspektive motiviert werden. Die Evaluation des öffentlichen Ressourceneinsatzes ist eine notwendige Voraussetzung für eine effektive und effiziente Staatstätigkeit im Sinne evidenzbasierter Politik (Cornet/ Webbink 2004).

Grundsätzlich ist die staatliche Intervention der FTI-Politik aufgrund einer soliden theoretischen Fundierung durch Marktversagen und Systemversagen legitimiert (Lundvall/ Borrás 2005). Trotzdem ist die Effektivität von FTI-Instrumenten keineswegs selbstverständlich, sondern letztlich eine empirisch zu klärende Problemstellung. Es mag etwa theoretisch gerechtfertigt sein, private F&E-Projekte durch öffentliche Förderung aufgrund positiver externer Effekte zu subventionieren, aber ob die geförderten Projekte tatsächlich Spillovers in einem bestimmten Ausmaß generieren, wäre z.B. empirisch zu untersuchen (Link/ Scott 2011). Weiterhin ist zu bedenken, dass die Entwicklung des Wirtschafts- und Innovationssystems eine Dynamik erzeugt, die eine ständige Überprüfung der Adäquatheit des aktuellen Politikmix notwendig machen. So machte etwa die Systemevaluierung zahlreiche Vorschläge zur Anpassung des Instrumenteneinsatzes an die geänderten Rahmenbedingungen durch den weitgehenden Abschluss des ökonomischen und technologischen Aufholprozess Österreichs an führende OECD-Staaten (Aiginger et al. 2009).

Während dies zu allen Zeitpunkten wünschenswert erscheint und eine anhaltend hohe Staatsquote in Österreich dieses Argument unterstreicht, ergeben sich durch die seit 2008 andauernde globale Wirtschaftskrise zusätzliche Notwendigkeiten in mehrerlei Hinsicht: Erstens sind öffentlichen Haushalte aufgrund der stark zugenommenen Staatsschulden im Sinne eines trendhaften Schuldenabbaus zu sanieren. Dies erfordert die Reduktion, Streichung oder Neuausrichtung von ineffektiven bzw. relativ ineffizienten Politikmaßnahmen. Das WIFO empfiehlt im Rahmen einer „Spitzenreiterstrategie unter Budgetknappheit“ die Steigerung der Hebelwirkung sowie der Effektivität und Effizienz des Fördersystems (Janger et al. 2010).

Zweitens sind aber auch neue Wachstumsquellen dringend gesucht. Diese sollen zum einen ein „Herauswachsen“ aus der Schuldenkrise ermöglichen und zum anderen einen Strukturwandel in Richtung eines nachhaltigeren ökonomischen Systems erbringen. Einige der relevanten Wachstumsfaktoren können und sollen von der Politik gefördert werden (OECD 2010).

Dies würde eine selektive Expansion der Mittel jener Förderinstrumente nahelegen, die in der Lage sind, entsprechende Wachstumsimpulse zu induzieren. Eine Identifikation dieser Instrumente kann letztlich nur auf empirischem Wege erfolgen. In diesem Sinne ist beispielsweise die Anhebung der Forschungsprämie von 8% auf 10% nicht unproblematisch, da dies ohne Evaluierung derselben (aufgrund mangelnder Datenbasis) erfolgt ist. Immerhin kostet die Anhebung der Forschungsprämie von acht auf zehn Prozent laut Budgetbericht rund 80 Mio. Euro im Jahr 2012. Auch die zweite, kürzlich gesetzte Maßnahme der Bundesregierung zur Erreichung einer F&E-Quote von 3,76% bis 2020, die Anhebung des Innovationschecks von 5000 € auf 10.000 €, erfolgte ohne Abwarten der Evaluierungsergebnisse des alten Innovationschecks.

Zusätzlich zu dieser krisenbedingten Ausnahmesituation werden mit dem Themenmanagement sowie der „Wirkungsorientierten Haushaltsführung“ (WHF) zwei institutionelle Neuerungen mit mittelbaren bzw. mittelbaren Auswirkungen auf die Evaluierung von FTI-Politik aktuell diskutiert und implementiert. Dies soll im Rahmen dieser Studie reflektiert und im Hinblick auf die Evaluation von FTI-Politik analysiert werden. Neben diesen institutionellen Aspekten hat diese Studie vor allem aber eine methodische Ausrichtung. Dabei geht es nicht nur um die Darstellung einiger, neuerer innovativer Methoden, sondern auch um einen nationalen und internationalen Vergleich der Methodenanwendung bei FTI-Evaluierungen sowie um eine Diskussion von Faktoren, welche als handlungsleitend bei Selektion einer passenden Evaluierungsmethode behilflich sein sollen. Hierbei gilt es der Frage nach der Datenverfügbarkeit besondere Aufmerksamkeit zu widmen.

Betrachtet man zunächst die methodischen Inhalte, so ist festzustellen, dass bereits mehrere einschlägige Studien existieren, welche in Form von Methodenüberblicken das methodische Instrumentarium von F&E-Evaluierungen darstellen. Nachstehende Box 1 zeigt die wesentlichen Publikationen im Überblick. Nachdem die hier vorliegende Studie eine Differenzierung zu den bisherigen einschlägigen methodischen Arbeiten anstrebt, sind Letztere als Komplementär zu den Ausführungen in dieser Studie zu verstehen.

Box 1: Kommentierter Methodenüberblicke zur Evaluation von FTI-Maßnahmen

Georgiou, L.; Roesner, D. (2000): Evaluating technology programs: tools and methods. Research Policy, 4-5, 657-678.

Klassischer Aufsatz der nicht von Methoden ausgeht, sondern unterschiedliche Instrumente zum Ausgangspunkt nimmt um Möglichkeiten der Evaluation darzustellen. Durch die Analyse von konkreten Fallbeispielen vermittelt das Paper auch wertvolle Einblicke in die Praxis der Methodenanwendung. Der Fokus liegt auf interviewbasierten Evaluierungsmethoden. Das Anspruchsniveau ist nicht zuletzt aufgrund der rein verbalen Darstellungsform eher niedrig.

Fahrenkrog, G. et al. (2002): RTD Evaluation Toolbox. Assessing the socio-economic impact of RTD-policies

Hierbei handelt es sich um eine sehr umfangreiche Methodensammlung (ca. 300 Seiten!). Es wird so gut wie jede in Frage kommende Methode zusammen mit ihren jeweiligen Möglichkeiten und Grenzen dargestellt. Die Methoden umfassen einen Mix aus qualitativen und quantitativen Methoden. Fallweise werden auch algebraische Darstellungsformen präsentiert.

Czarnitzki, D. et al. (2003): Einsatzmöglichkeiten neuer quantitativer Ansätze zur Wirkungsanalyse öffentlicher Forschungsförderung – Eine Grundlagenstudie. Methodische Grundlagen, Bestandsaufnahme und Anwendung am Beispiel Mikrosystemtechnik.

Dirk Czarnitzki ist einer der führenden und produktivsten Innovationsökonomien in Europa, der sich ausführlich mit Fragen der quantitativ-analytischen Evaluierung von FTI-Maßnahmen beschäftigt. Diese Studie des Zentrums für Europäische Wirtschaftsforschung ist ein Plädoyer für ökonometrische Evaluationsmethoden und zeigt deren Potenzial an einem konkreten Anwendungsfall. Nachdem die Methoden zu einem Gutteil algebraisch präsentiert werden, ist dieser Überblick nur für geübte Leser mit mathematischer und statistischer Vorkenntnis zu empfehlen.

Ruegg, Feller (2003) A toolkit for evaluating public R&D investment. Models, methods and findings from ATP's first decade.

Diese Publikation vermittelt einen Eindruck über FTI-Evaluationsmethodeinsatz in den USA am Beispiel des 2007 eingestellten Advanced Technology Programs. Etwas anders als in Europa steht der Einsatz von FTI-Maßnahmen in den USA unter stetigem kritischem Vorbehalt. Entsprechend sind die Förderagenturen, nicht zuletzt auch durch verschiedene einschlägige Gesetze, dazu gezwungen, ihre Aktivitäten zu rechtfertigen. Die angewandten Methoden unterschieden sich zum Teil von jenen in Europa. Die präsentierten Methoden umfassen hier jedoch den kompletten Kanon und sind insofern mit der RTD-Evaluation Toolbox vergleichbar. Das Niveau der Darstellung ist jedoch etwas einfacher gehalten.

Rhomberg et al. (2006): Neue Entwicklungen im Bereich der Wirkungsanalyse und -abschätzung FTI-politischer Maßnahmen.

Diese vom heutigen Austrian Institute of Technology durchgeführte Studie wurde vom Rat für Forschung und Technologieentwicklung beauftragt. Hintergrund dieser Beauftragung war neben dem generellen Interesse des RFTE an Fragen der Evaluierung insbesondere die Abschätzung der Möglichkeiten zur Wirkungsanalyse der vom RFTE empfohlenen Offensivprogramme. Inhaltlich ähnelt diese Studie der RTD-Evaluation Toolbox, wobei das Anspruchsniveau der Darstellung geringer ist bei etwa gleicher inhaltlicher Breite. Zusätzlich wird der Status quo österreichischer Wirkungsanalysen präsentiert.

Um bei dieser bereits umfangreichen Literaturlage einen Mehrwert zu leisten, werden im methodischen Teil insbesondere folgende Aspekte herausgearbeitet: (1) Anstatt eines weiteren, umfassenden Überblicks wird eine empirisch und theoretisch begründete Auswahl aus dem Methodenspektrum dargestellt. (2) Berücksichtigung der spezifisch österreichischen Evaluierungssituation hinsichtlich Datenverfügbarkeit und Methodeneinsatz im internationalen Vergleich. (3) Darstellung methodischer Weiterentwicklungen bzw. neuer Schwerpunktsetzungen

seit der Publikation der vorläufig letzten einschlägigen Studie von Rhomberg et al. (2006). Trotzdem kann eine gewisse Überschneidung mit anderen Studien freilich nicht vermieden werden. Im Sinne des zeitlichen Abstandes ist aber auch die Replikation oder Falsifikation älterer Studienergebnisse ein durchaus interessantes Ergebnis.

Wie bereits erwähnt ist das übergeordnete Ziel der Studie die Analyse der Grenzen und Möglichkeiten in der Anwendung wirkungsorientierter Evaluation. Dieses Ziel soll unter Berücksichtigung des Förderportfolios der FFG und der thematischen Schwerpunktsetzungen des BMVIT verfolgt werden. Zu diesem Zweck werden in dieser Studie insbesondere adäquate Evaluationsmethoden für FTI-bezogene Förderprogramme des BMVIT und der FFG vorgestellt. Diese übergeordnete Zielsetzung wird anhand von fünf, eng miteinander verknüpften Teilzielen operationalisiert:

1. Analyse und Vergleich nationaler und internationaler Evaluierungstrends von FTI-Maßnahmen
2. Darstellung potenzieller Auswirkungen der Umstellung von programmorientierter Förderung hin zu einem Fördermanagement entlang thematischer Leitlinien („Themenmanagement“) in der FFG sowie der wirkungsorientierten Haushaltsführung auf die Evaluierungspraxis im FTI-Bereich
3. Aufzeigen innovativer Evaluierungsmethoden zur Förderung von Kreativität und Aussagekraft von Evaluierungen
4. Förderung realistischer Erwartungsbildung auf Seiten der Evaluierungsnachfrager über Möglichkeiten und Grenzen unterschiedlicher Evaluierungsmethoden
5. Ableitung von Empfehlungen zur weiteren Entwicklung eines effektiven Evaluierungssystems im österreichischen FTI-Sektor mit Schwerpunkt FFG und BMVIT.

Zur Erreichung dieser Ziele wurden unterschiedliche Methoden eingesetzt. Dabei handelt es sich einerseits um qualitative Interviews mit insgesamt 11 Stakeholdern von FFG, BMVIT, BMF sowie diversen Evaluierungsanbietern (siehe Anhang B). Im Rahmen von Dokumentenanalysen wurden die österreichischen FTI-Evaluierungen der letzten Jahre und Programmdokumente analysiert. Der internationale Vergleich von eingesetzten Evaluierungsmethoden erfolgte über eine Internetrecherche bei ausgewählten Förderagenturen sowie eine Literaturrecherche in mehreren Fachzeitschriften.

Die Studie ist inhaltlich wie folgt gegliedert. Kapitel 2 spannt zunächst den Kontext für FTI-Evaluationen unternehmensbezogener F&E-Programme in Österreich auf. Dabei wird neben einem kurzen Überblick über relevante Institutionen vor allem die Höhe und Struktur der F&E-Subventionen dargestellt. Eine kurze Präsentation des FFG-Programmportfolios sowie thematischer Schwerpunktsetzungen des BMVIT runden dieses Kapitel ab.

Kapitel 3 gibt eine Darstellung zum Status quo österreichischer FTI-Evaluierungen anhand der Interviewergebnisse und setzt diese in Bezug zu älteren Studien und anderen Literaturquellen. Von aktueller Bedeutung ist dabei insbesondere der Abschnitt 3.2. Hier werden das Themenmanagement als neue Förderstrategie der FFG und die Wirkungsorientierte Haushaltsführung als neue Budgetstrategie des Bundes in ihren Grundzügen und Implikationen für FTI-Evaluierungen erörtert.

Kapitel 4 stellt nationale und internationale Evaluierungstrends von FTI-Evaluationen mit einem besonderen Fokus auf methodische Aspekte dar. Dabei bestätigt sich eindrucksvoll der bereits bekannte Befund eines relativen Mangels von ökonometrischen FTI-Evaluierungen in Österreich relativ zu anderen OECD-Staaten. Als methodische Spezialisierung des angelsächsischen Raums wird hier auch die Kosten-Nutzen Analyse erörtert.

Kapitel 0 bildet das Methodenkapitel, wobei dieses in einen qualitativen und einen quantitativen Teil untergliedert ist. Zum Abschluss erfolgt eine kurze Reflexion über die Frage, welche Fragestellungen bzw. Instrumentencharakteristika den Einsatz welcher Methode verlangen bzw. nahelegen.

Kapitel 0 diskutiert die in letzter Zeit zunehmend wichtiger werdenden sozioökonomischen Zielsetzungen von FTI-Programmen. Dabei stellt sich vor allem das Problem, wie derart umfassende Ziele operationalisiert und evaluiert werden können.

Kapitel 7 diskutiert einen weiteren aktuellen und kritischen Aspekt von FTI-Evaluation, der einen besonderen Stellenwert für Österreich besitzt. Es handelt sich hierbei um die Frage der Datenverfügbarkeit. Dieses Thema wird vor allem auch durch einen internationalen Vergleich interessant, weil sich hier signifikante Unterschiede zwischen den Staaten zeigen. Demnach könnte Österreich einiges von anderen Staaten über einen effizienten Umgang mit amtlichen Daten für öffentliche Zwecke lernen.

Ein Resümee und politische Schlussfolgerungen fassen die Studie zusammen und leiten Handlungsempfehlungen aus den Studienergebnissen für das politische System ab.

2 Innovationspolitischer Kontext von FTI-Evaluierungen in Österreich

„Die Konkurrenz der Ministerien auf dem Gebiet der FTI-Politik hat zu einem Wettlauf um die Erfindung von Programmen der direkten Förderung geführt, zu dem viel diskutierten ‚Wildwuchs‘, der so gut wie alle (kurzfristigen) Bedürfnisse abdeckt und die inkrementale Innovation kräftig fördert, auch wenn fast jedes Programm zwangsläufig unterdotiert ist. In manchen Fällen stehen hinter diesen Programmen effiziente Lobbys, vielfach aber initiative und durchsetzungskräftige BeamtInnen.“

Tichy (2009:269)

Wie auch in anderen Ländern hat das Politikfeld Innovationspolitik in den vergangenen 10 Jahren an Bedeutung gewonnen (Zinöcker/ Dinges 2009).¹ Insbesondere die Lissabon-Agenda und das damit verbundene Barcelona-Ziel einer F&E-Quote in der Höhe von 3% sorgten für die Mobilisierung zusätzlicher privater und öffentlicher Ressourcen sowie für die Internationalisierung und weitere Professionalisierung der FTI-Politik. Immerhin wurden 2010 öffentliche und private Ressourcen in der Höhe von insgesamt 7,89 Mrd. € für F&E ausgegeben, das sind ca. 2,78% des BIP. Davon wurden 38,9% vom öffentlichen Sektor (Bund, Länder, Gemeinden) finanziert.

Abschnitt 2.1 skizziert den institutionellen Kontext von FTI-Evaluierungen in Österreich und diskutiert rezente Stellungnahmen FTI-politischer Akteure zur Funktion und Bedeutung von Evaluation. Danach wird in Unterkapitel 2.2 die Struktur und Entwicklung der österreichischen F&E-Förderungen im internationalen Vergleich erörtert. Gemäß den Schwerpunktsetzungen dieser Studie werden im Anschluss daran das Programmportfolio der FFG sowie die thematischen Schwerpunktsetzungen des BMVIT vorgestellt.

2.1. INSTITUTIONEN, AKTEURE UND STRATEGIEN

Institutionell wird das Politikfeld auf Ministeriumsebene durch die drei Ministerien (1) Bundesministerium für Verkehr, Innovation und Technologie (BMVIT), (2) Bundesministerium für Wissenschaft und Forschung (BMWFi), und das (3) Bundesministerium für Wirtschaft, Familie und Jugend (BMWFiJ) bestimmt. Diese sind auch die wesentlichen Auftraggeber für Evaluierungen auf Bundesebene. Die Gegenstände der Evaluierung, d.h. insbesondere FTI-Programme, werden von dafür gegründeten Agenturen bzw. Projektträgern durchgeführt. Im Bereich der angewandten Forschung sind dies insbesondere die Forschungsförderungsgesellschaft (FFG) sowie das Austrian Wirtschaftsservice (AWS). Die Förderung der Grundlagenforschung erfolgt vorrangig durch den Fonds zur Förderung der wissenschaftlichen Forschung (FWF). Während der FWF selbst als aktiver Nachfrager nach Evaluierungen auftritt, hat die FFG in den letzten Jahren zunehmend weniger eine

¹ Eine umfassendere Darstellung zum Kontext österreichischer FTI-Evaluierungen gibt der von der Plattform für Forschungs- und Technologieevaluierung herausgegebene Sammelband „Evaluation of Austrian Research and Technology Policies. A Summary of Austrian Evaluation Studies from 2003 to 2007“. Für einen Vergleich mit Deutschland siehe Kuhlmann (2003).

eigenständige Evaluierungsnachfrage gezeigt. Stattdessen kommt es vermehrt zur Vergabe von Studien zur Orientierung über spezifische Aspekte des österreichischen Innovationssystems.

Damit kommt die Nachfrage nach Evaluierungen der FFG vorrangig von den beiden Trägern der FFG, dem BMVIT und dem BMWFJ.

Wie auch in anderen Ländern hat auch das Finanzministerium zunehmend an Einfluss relativ gegenüber den anderen Ministerien gewonnen (Klump 2006). Dies gilt auch für den FTI-Bereich, der durch abnehmenden Einfluss des Bundeskanzleramts und zunehmenden Einfluss des Finanzministeriums gekennzeichnet ist (Gottweis/ Latzer 2006). Eine Funktion von Evaluierungen aus Sicht der beauftragenden Ministerien ist unter anderem die Legitimation der Mittelverwendung gegenüber dem BMF, die etwa bei der Frage der Weiterführung eines Programms in Form von Evaluierungen geleistet werden kann. Während der unmittelbare Einfluss des BMF bislang primär über den Prozess der Budgetplanung und -erstellung und bei der Begutachtung im Rahmen des Designs neuer FTI-Programme realisiert wurde, kommt als zusätzliches Instrument ab 2013 die Einführung einer sog. Wirkungsorientierten Haushaltsführung hinzu. Eine etwas untergeordnete Rolle spielt das BMF auch bei der Vergabe von Evaluierungen, wenn MitarbeiterInnen des BMF als Jurymitglieder mitwirken.

Auf der Anbieterseite von Evaluation besteht ebenso wie auf der Nachfrageseite eine oligopolistische Marktstruktur.² Während die Universitäten kaum Aktivitäten im Bereich Evaluierungen setzen, wird das Feld durch einige außeruniversitäre Forschungsinstitute und wenige private Anbieter dominiert (JOANNEUM RESEARCH, KMU-Forschung Austria, AIT, eingeschränkt WIFO, Technopolis, GFF, ZSI, Convelop). Sprachliche Barrieren sowie die Notwendigkeit detaillierter institutioneller Kenntnisse tragen zu einer relativ starken Marktsegmentierung entlang nationalstaatlicher Grenzen bei.

Weitere relevante Akteure für FTI-Evaluierungen sind der Rat für Forschungs- und Technologieentwicklung (RFTE) sowie die Plattform Forschungs- und Technologieevaluierung (fteval), die, in neokorporatistischer Tradition, zur Etablierung gemeinsamer Standards von FTI-Evaluierungen beiträgt (Pichler 2009). Last but not least ist auch noch der Rechnungshof zu erwähnen, der im Rahmen seiner gesetzlichen Prüfpflicht über die sparsame, wirtschaftliche und zweckmäßige öffentliche Mittelverwendung auch Instrumente der FTI-Politik überprüft (z.B. Rechnungshof, Bund 2009/9).

Nach dieser Darstellung des innovationspolitischen Kontexts stellt sich die Frage, welche Erwartungen bzw. Vorstellungen die österreichische FTI-Politik an bzw. von Evaluation hat? Zur Beantwortung dieser Frage bietet sich ein Blick in aktuelle und strategisch bedeutsame Policy-Dokumente an.

² Es sei an dieser Stelle darauf hingewiesen, dass die Marktform des Monopols nicht zwangsläufig mangelnden Wettbewerb bedeutet. Im Oligopolmodell nach Bertrand kommt es sogar zu Marktergebnissen, die denjenigen bei vollkommener Konkurrenz entsprechen. Dieses sogenannte Bertrand-Paradoxon trifft insbesondere dann zu, wenn die gehandelten Güter einen hohen Homogenitätsgrad aufweisen. Nach Aussagen mancher Interviewpartner trifft dies für Evaluierungsangebote jedenfalls teilweise zu.

Box 2: Aktuelle Stellungnahmen zur Evaluierung von FTI-Maßnahmen

OECD Innovation Strategy (2010:211)

„Evaluation of policies is essential to enhance the effectiveness and efficiency of policies to foster innovation. Effective evaluation is also the key for the legitimacy and credibility of government intervention in innovation processes. Improved approaches and methods for evaluation are required [Herv. d. Verf.] to capture the broadening of innovation as is better feedback of evaluation into the policy making process.“

Strategie 2020 des Rats für FTE (2010:44)

„Methoden und Modelle zur Analyse des FTI-Systems und für die Wirkungsforschung sind weiterzuentwickeln und wissenschaftliche Expertise dazu aufzubauen [Herv. d. Verf.]

Forschungsrelevante Entwicklungen in anderen Bereichen (z.B. in Gesamtwirtschaft und Hochschulsektor sollen einbezogen werden.“ [Anm. d. Verf. Dieses Ziel soll bis 2013 erreicht werden].

FTI-Strategie des Bundes (2011:11)

„Bei allen Maßnahmen steht die Wirkungsorientierung im Vordergrund. Initiativen und Programme, die nicht zu den gewünschten Wirkungen führen, werden eingestellt oder grundlegend neu ausgerichtet [Herv. d. Verf.]“

Alleine ein cursorisches Lesen dieser Zitate zeigt, dass hier offenbar ziemlich ambitionierte Ziele und Erwartungen formuliert werden. Sowohl die OECD als auch der RFTE weisen auf die Notwendigkeit einer Weiterentwicklung von Evaluationsmethoden hin. Es wird der Aufbau einer „wissenschaftlichen Expertise“ gefordert – und das bis 2013. Es darf an dieser Stelle gefragt werden, wie und vor allem wer derartige Grundlagenforschungsaktivitäten finanzieren soll. Bislang waren die durchschnittlichen Evaluationsbudgets jedenfalls weitgehend unzureichend, um gleichsam als Nebenprodukt Methodenforschung zu betreiben. Die FTI-Strategie des Bundes weist Evaluierungen mit der Perspektive der Einstellung von ineffektiven Programmen einen potenziell sehr hohen Einfluss zu. Die Erhöhung der Forschungsprämie zeigt jedoch, dass die Selbstbindung an diese Forderungen bislang unzureichend ist.

2.2. F&E-FÖRDERUNGEN IM ÜBERBLICK

Österreich gibt relativ viel Geld für die Förderung unterschiedlichster Aktivitäten aus. Genauer gesagt sogar relativ am meisten im Vergleich zu allen anderen Staaten in der EU. Wenngleich festgestellt werden muss, dass die statistischen Angaben sowie die unterschiedlichen Verfahren zur Aggregation von Subventionszahlungen zu erstaunlich unterschiedlichen Ergebnissen kommen (Buigues/ Sekkat 2011). Tabelle 1 zeigt Berechnungen aus einer aktuellen Studie, wobei Österreich im Durchschnitt der Jahre 1998-2002 den höchsten Anteil von Subventionszahlungen am BIP aufwies. Auffallend ist auch der relativ große Abstand zu Dänemark, dem Staat mit den zweithöchsten Subventionen. Statistisch betrachtet ist Österreich dabei ein positiver Ausreißer in der Verteilung. Interessant ist auch, dass Österreich als einziges Land ein positives Wachstum bei Subventionen zwischen den Durchschnittswerten der Perioden 1975-1990 und 1998-2002.

Das ist ein weithin konterintuitives Ergebnis, würde man doch aufgrund von Sparprogrammen seit den 1980er Jahren und aufgrund des EU-Beitritts eine deutliche Reduktion der Subventionszahlungen erwarten.

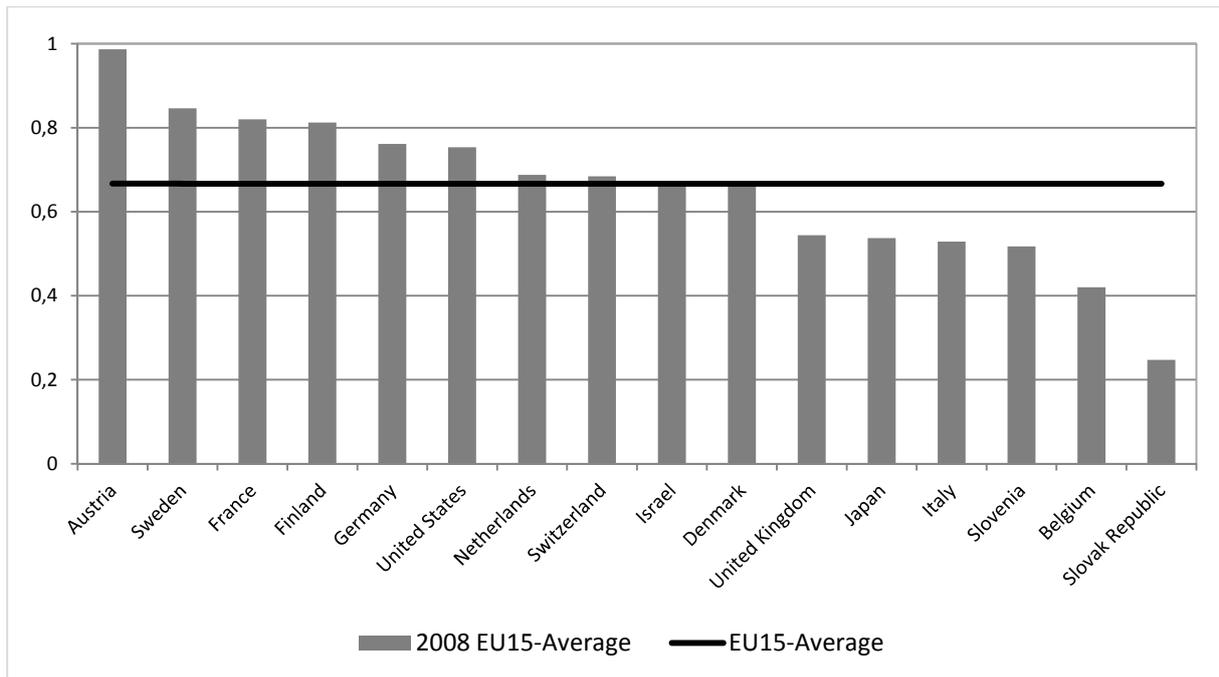
Tabelle 1: Niveau und Trends staatlicher Subventionen im internationalen Vergleich (in % des BIP)

	Average 1998–2002	Average 1975–1990	Difference
Austria	3,0	2,9	+0,1
Denmark	2,2	3,2	-1,0
Sweden	1,8	4,5	-2,7
Germany	1,7	2,1	-0,4
Finland	1,5	3,2	-1,7
Belgium	1,5	3,9	-2,4
Netherlands	1,5	2,9	-1,4
Portugal	1,4	3,7	-2,3
France	1,3	2,7	-1,4
Italy	1,2	3,3	-2,1
Spain	1,1	2,1	-1,0
Ireland	0,8	7,5	-6,7
Japan	0,8	1,2	-0,4
united Kingdom	0,5	2,1	-1,6
United States	0,5	0,5	0
Greece	0,2	4,2	-4,0

Buigues/ Sekkat 2011

Zeigt das Niveau der aggregierten Subventionen bereits einen Spitzenwert für Österreich, so setzt sich dieses Ergebnis auch in den entsprechenden Datensätzen zur staatlichen Subvention von F&E-Aktivitäten fort. Österreich weist den höchsten Subventionsanteil der gesamten volkswirtschaftlichen F&E-Ausgaben auf (Abbildung 1). Nachdem diese Studie schwerpunktmäßig den Förderungen von BMVIT und FFG gewidmet ist, wird nachstehend nur auf unternehmensbezogene Förderungen näher eingegangen.

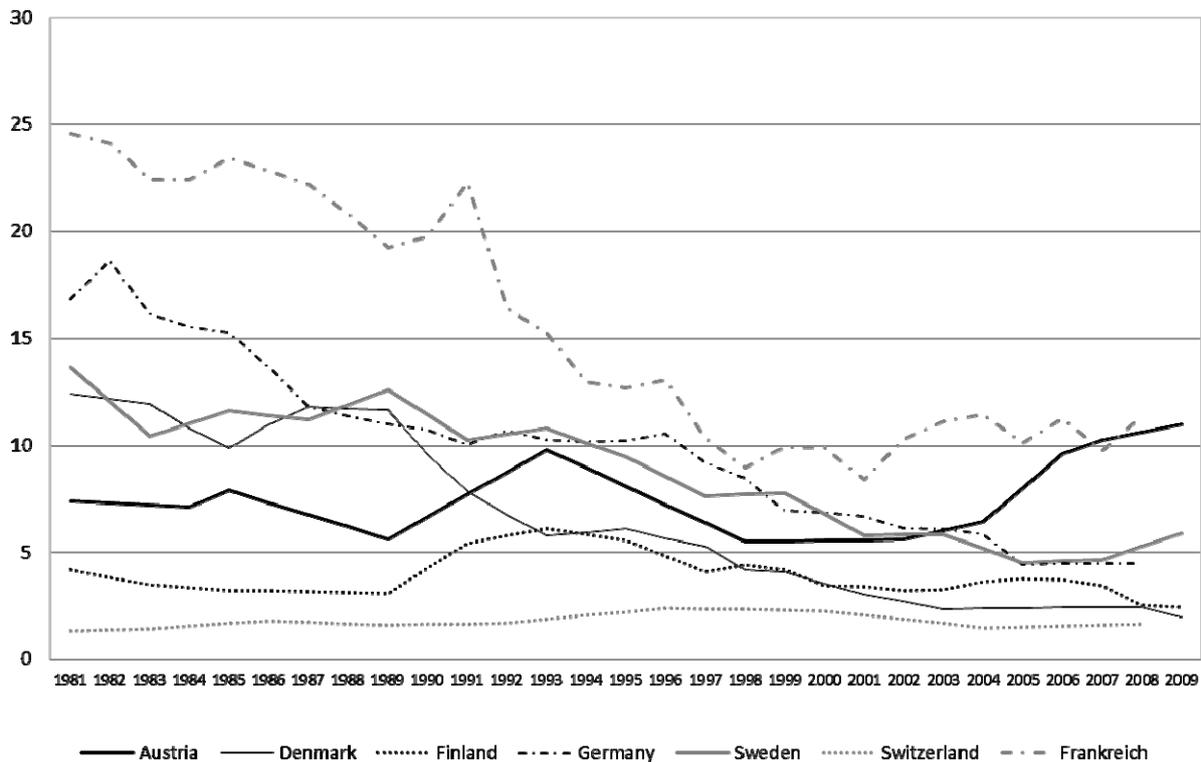
Abbildung 1: Anteil des Staates an den volkswirtschaftlichen F&E-Ausgaben (GERD) in % des BIP (2008)



OECD, MSTI. Notes: The values from Sweden, Netherlands, Denmark and Belgium are from 2007.

Abbildung 2 stellt die Entwicklung des staatlichen Finanzierungsanteils an den F&E-Ausgaben des Unternehmenssektors für die letzten drei Jahrzehnte für Österreich und ausgewählte europäische Vergleichsländer dar. Im Einklang mit den oben formulierten Erwartungen über die Entwicklung von Subventionszahlungen zeigt sich ein generell fallender bzw. stagnierender öffentlicher Finanzierungsanteil privater F&E-Tätigkeiten. So fielen diese Anteile etwa in Frankreich von etwa 25% im Jahr 1981 auf nunmehr ca. 10% für das Jahr 2009 und in Deutschland von ca. 12% auf unter 5%. Auch Dänemark verzeichnete eine starke Abnahme; dies passt auch gut zu den Ergebnissen in Tabelle 1. Auf der anderen Seite zeigen „Innovation Leaders“ wie Finnland oder die Schweiz konstant niedrige Subventionszahlungen für unternehmerische F&E-Aktivitäten. Im Unterschied zu den Vergleichsstaaten nahm der Förderanteil in Österreich im letzten Jahrzehnt, ausgehend von einem mittleren Niveau, von ca. 5% auf über 10% stark zu. Das entspricht einer Verdopplung innerhalb von weniger als 10 Jahren. Damit fördert neben Österreich nur noch Frankreich in einem ähnlich hohem Ausmaß unternehmerische F&E. Allerdings muss bei dieser Statistik bedacht werden, dass das Niveau der BERD in Ländern wie Schweden (2,5%, 2009) oder Finnland (2,8%, 2009) wesentlich höher liegt als in Österreich (1,9%, 2009).

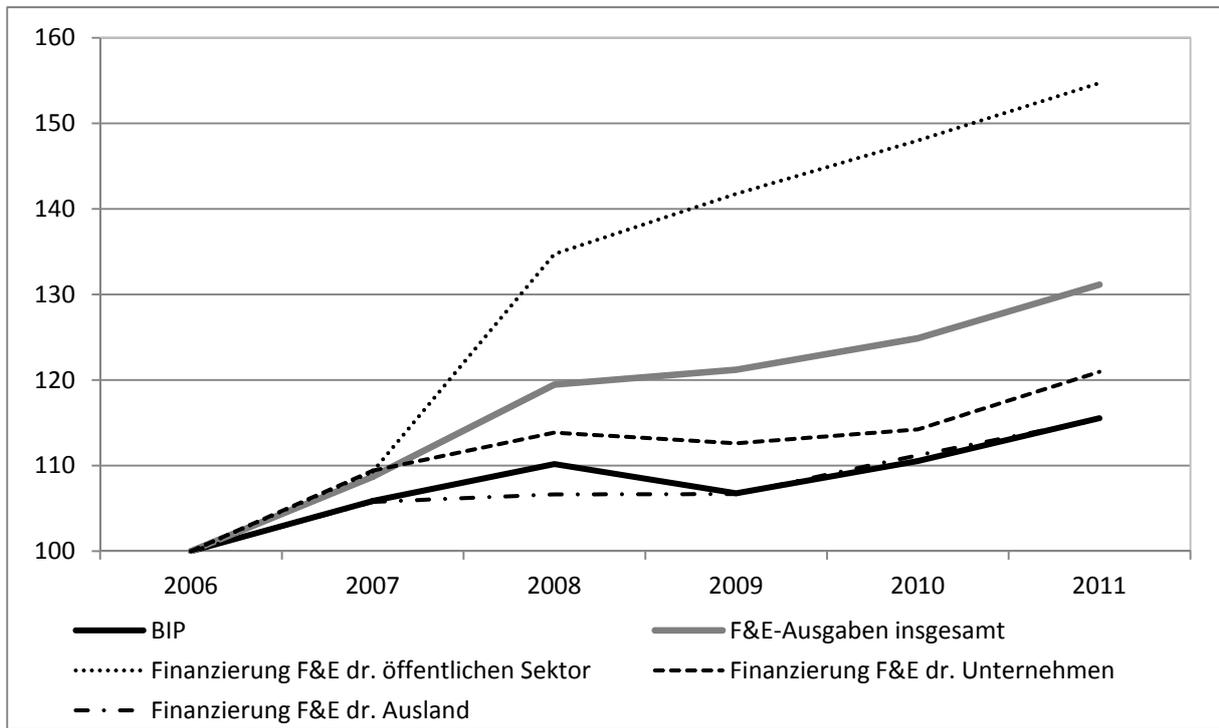
Abbildung 2: Prozentualer Anteil der vom Staat finanzierten F&E-Ausgaben des Unternehmenssektors



Eigene Darstellung, Daten: OECD MSTI, lineare Interpolation fehlender Werte

Bemerkenswert ist die Entwicklung der österreichischen F&E-Ausgaben auch in der Wirtschafts- und Finanzkrise, die ab 2007, ausgehend von den USA, eine globale Rezession verursachte. Abbildung 3 zeigt die Entwicklung des BIP sowie der relevanten Aggregate der F&E-Erhebung. Zunächst einmal kann festgestellt werden, dass, im Gegensatz zur Entwicklung des BIP, die Gesamtausgaben für F&E keine Abnahme über den Zeitraum der Krise aufweisen. Der Wachstumstrend ist für das Jahr 2008 zwar wesentlich abgeschwächt, aber er bleibt leicht positiv. Wesentliche Ursache hierfür ist die deutliche Zunahme der F&E-Finanzierung durch den öffentlichen Sektor. Demgegenüber kam es im Unternehmenssektor zu einem leichten Rückgang im Jahr 2008 um ca. 40 Mio. Euro. Die Finanzierung durch das Ausland stagnierte in diesem Jahr. Damit hat der öffentliche Sektor einen deutlich antizyklischen Impuls bei den F&E-Ausgaben gesetzt und eine Abnahme der volkswirtschaftlichen F&E-Ausgaben verhindert.

Abbildung 3: Entwicklung von BIP und F&E-Ausgaben 2006-2011 (2006=100)



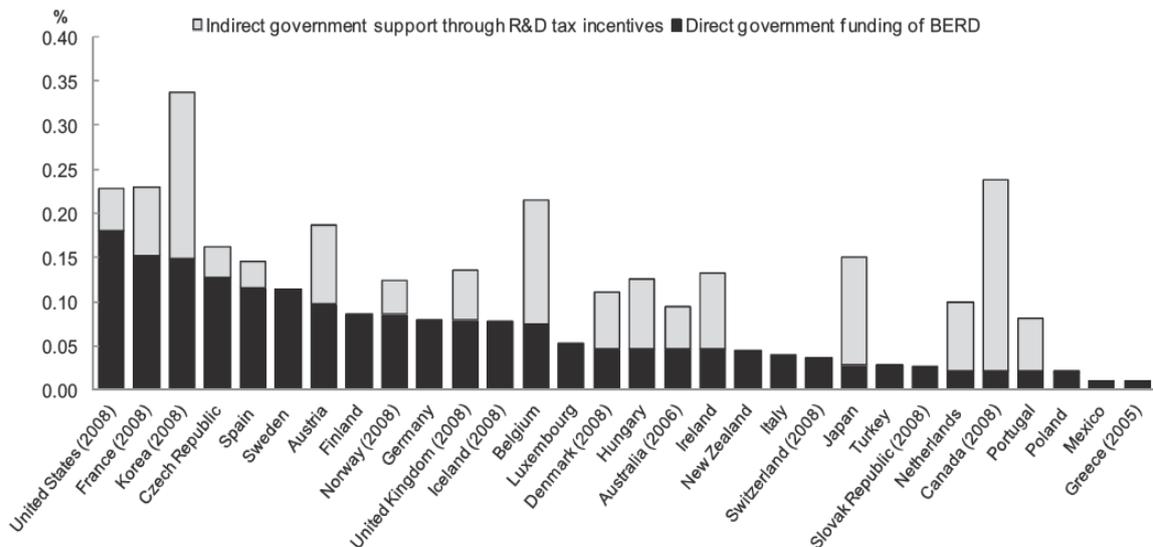
Eigene Darstellung, Daten: Statistik Austria. Der öffentliche Sektor ist eine Aggregation von Bund, Bundesländer und „Sonstige“, wobei hierunter u.a. die Gemeinden, Kammern sowie die Sozialversicherungen zählen.

Wie stellt sich nun die Struktur der Förderungen im Sinne von direkter vs. indirekter F&E-Förderung dar? Abbildung 4 zeigt eine Verteilung von ca. 50:50 zwischen direkter und indirekter Förderung. Durch die Erhöhung der Forschungsprämie – die hier noch nicht berücksichtigt ist – dürfte sich dieses Verhältnis aber entsprechend verschieben. Vergleicht man nur die staatlichen Ausgaben für direkte Forschungsförderung, so zeigt sich, dass diese in Österreich ebenfalls im internationalen Vergleich zu den höchsten zählen. Gleichzeitig ist aber auch festzustellen, dass z.B. Schweden, mit einem insgesamt nur etwa halb so hohen Subventionsniveau unternehmerischer Innovationsaktivitäten, höhere Ausgaben für direkte Forschungsförderung als Österreich tätigt. Generell zeigt sich, dass die OECD-Staaten sehr unterschiedliche Strategien hinsichtlich ihres Instrumentenmix verfolgen (Jaumotte/ Pain 2005). Während z.B. Deutschland, Finnland, Schweden oder die Schweiz ausschließlich direkte Förderungen als Instrumente einsetzen, weisen Dänemark aber v.a. auch Belgien relativ hohe indirekte Fördervolumina auf. Angesichts des mit ca. 50% ebenfalls relativ hohen Beitrags der indirekten Forschungsförderung in Österreich sei hier nochmals darauf hingewiesen, dass damit de facto die Hälfte (!) des öffentlichen Ressourceneinsatzes für F&E bislang nicht evaluiert wurde und auch in naher Zukunft aufgrund mangelnder Datenverfügbarkeit nicht evaluiert werden dürfte. Das WIFO verlangte in der Stellungnahme zur FTI-Strategie eine solche Evaluierung: „Für eine vollständige Effizienz- und

Effektivitätsbeurteilung sollte die Erreichung des zweiten Ziels der steuerlichen Förderung, der Intensivierung der F&E-Ausgaben, sorgfältig evaluiert werden. (..)

Eine Evaluierung wäre auch deshalb sinnvoll, weil die Forschungsprämie nicht im Bundesvoranschlag aufscheint und sich daher der wirkungsorientierten Haushaltsführung entzieht“ (Janger et al. 2010). Jaumotte und Pain (2005) weisen aber auch darauf hin, dass das Wissen über die Effekte indirekter Förderung generell schlechter ist als jenes über die Effekte direkter Förderung.

Abbildung 4: Direkte und indirekte Subvention von unternehmerischer F&E (BERD) 2008 (oder letztes verfügbares Jahr) in % des BIP

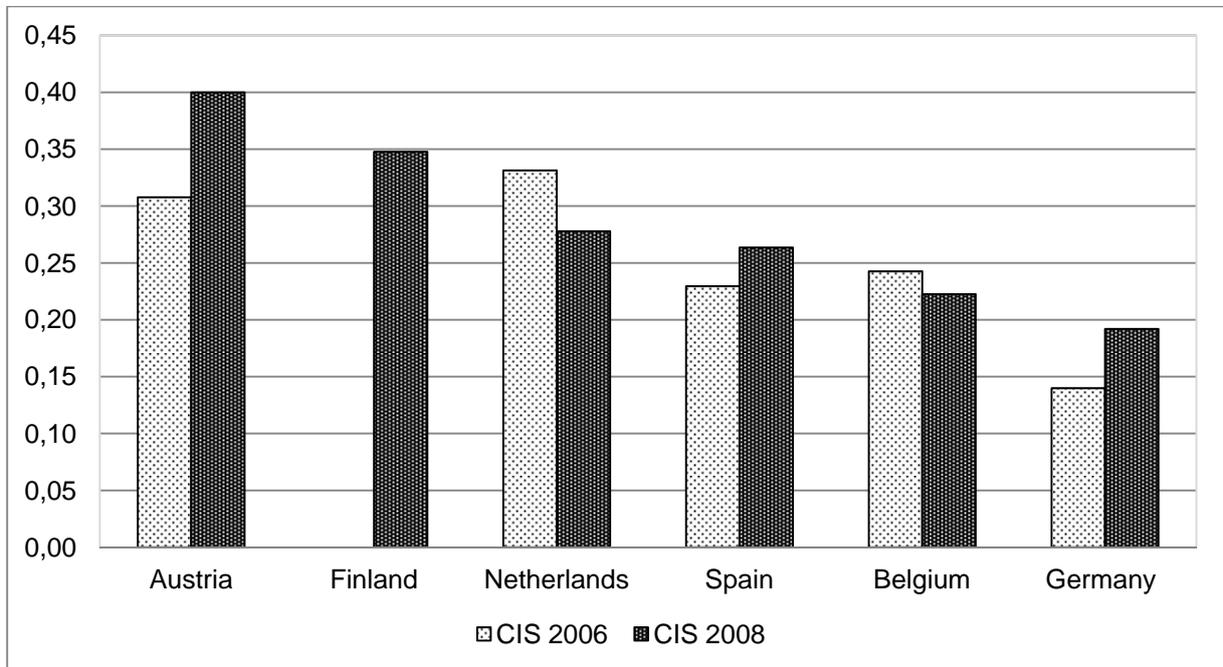


OECD 2010

Neben dem generellen Niveau der Subventionszahlungen für Innovationsaktivitäten stellt sich die Frage nach der Breite bzw. Streuung derselben. Freilich ist hier zunächst ein positiver Zusammenhang zwischen Subventionsniveau und dem Anteil von Unternehmen mit Innovationssubventionen zu erwarten, und tatsächlich zeigt Abbildung 5 auf Basis der Europäischen Innovationserhebung von 2008 den relativ höchsten Anteil an geförderten Unternehmen in Österreich. Interessant ist hier auch der Wert des Merkmals: Mit 40% erhält beinahe jedes zweite innovative Unternehmen in Österreich eine Innovationsförderung. Aber es scheint auch hier große Unterschiede in der „Förderbreite“ zwischen den Staaten zu geben, wengleich die Frage, ob diese tatsächlich intendiert sind, keinesfalls einfach beantwortet werden kann. So weist Österreich einen etwa doppelt so hohen Subventionsanteil an den BERD und einen beinahe doppelt so hohen Ressourceneinsatz in % des BIP zur Förderung von Unternehmensinnovationen im Vergleich zu Finnland auf, anteilmäßig werden jedoch in Finnland ebenfalls ca. 35% der innovativen Unternehmen gefördert. Interessant ist auch der Unterschied zwischen den Niederlanden und Spanien: Obwohl der öffentliche Sektor der Niederlande signifikant weniger Innovationsförderungen als Spanien finanziert, erhalten mehr Unternehmen eine Förderung. Generell lassen sich diese Unterschiede jedoch nicht einfach durch

unterschiedliche Fördersysteme im Sinne der Zusammensetzung von direkter und indirekter Förderung erklären.

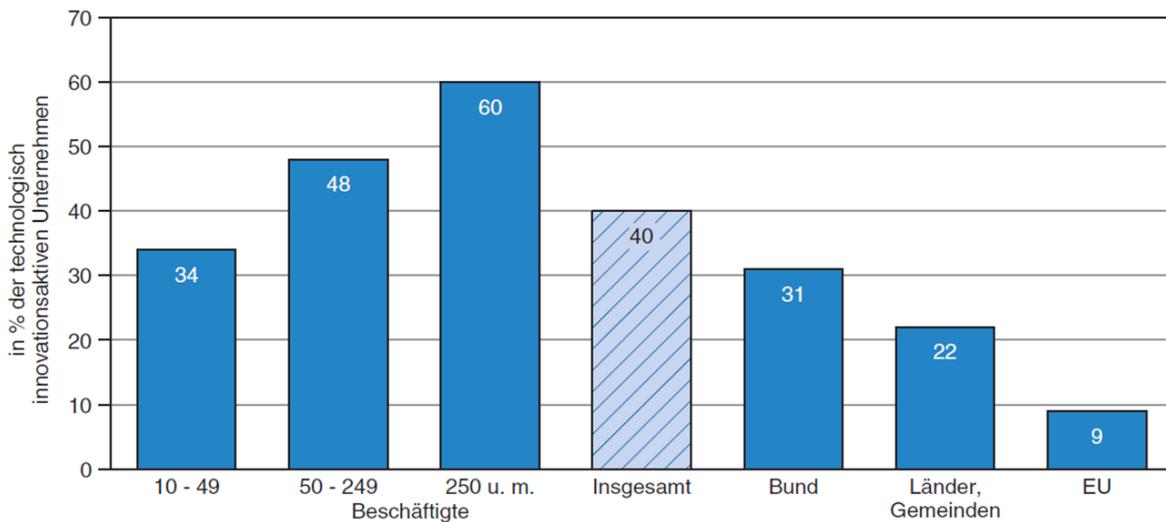
Abbildung 5: Anteil innovativer Unternehmen mit Innovationsförderung in % aller technologisch innovativen Unternehmen



Eigene Darstellung, Daten: Eurostat, CIS 2006, CIS 2008

Während Abbildung 5 einen Überblick über die Förderintensitäten des gesamten Unternehmenssektors gibt, zeigt Abbildung 6 den gleichen Indikator differenziert nach Unternehmensgrößenklasse und Fördergeber. Demnach besteht ein positiver Zusammenhang zwischen der Förderwahrscheinlichkeit innovativer Unternehmen und der Unternehmensgröße. So erhalten etwa 34% der innovativen Kleinunternehmen eine Förderung, während der entsprechende Wert bei Großunternehmen bei 60% liegt. Der Bund ist der wichtigste Fördergeber für Unternehmen, die EU spielt eine relativ geringe Rolle.

Abbildung 6: Innovative Unternehmen in Österreich mit Innovationsförderung differenziert nach Größenklasse und Fördergeber



Statistik Austria 2010 (CIS 2008)

Zwischen Datengrundlage und Evaluierungsmethoden besteht ein enger Zusammenhang (s. Kapitel 7). Für die Anwendung quantitativ-analytischer Methoden ist vor allem die absolute Anzahl geförderter Unternehmen von Bedeutung. Weiterhin ist zu bedenken, dass Gruppierungen der Unternehmen (z.B. nach Branchenzugehörigkeit) sowie geringe Rücklaufquoten bei Unternehmensbefragungen die Größe der jeweiligen Zellenbesetzung reduzieren. Wird eine Evaluierungsmethode gewählt, die Beobachtungen zu Nichtteilnehmern notwendig macht, so ist auch deren absolute Häufigkeit von Bedeutung. Tabelle 2 zeigt die entsprechenden absoluten Häufigkeiten. Interessant ist vor allem die Differenz zwischen den geförderten und nicht geförderten innovativen Unternehmen, da dieser Wert eine Abschätzung der Größenordnungen potenzieller Kontrollgruppen zulässt. Hier zeigt sich, dass die Bildung einer Kontrollgruppe für geförderte, innovative Großunternehmen Probleme bezüglich der Realisierung einer hinreichend großen absoluten Häufigkeit der Vergleichsgruppe bereiten könnte.

Tabelle 2: Absolute Häufigkeiten geförderter und nicht geförderter innovativer Unternehmen in Österreich (2008)

	Geförderte innovative Unternehmen ¹	Innovative Unternehmen (gesamt) ¹	Differenz geförderte Unternehmen-nicht geförderte Unternehmen	Unternehmen gesamt ²
10-49	1509	4426	2917	294.434
50-249	835	1723	888	5192
>249	353	587	234	1119

1: Innovativ ist hier im Sinne von technologischer Definition zu verstehen. Daten: CIS 2008.

2: Daten: Leistungs- und Strukturerhebung 2008

Zusammenfassend ergeben sich aus der deskriptiven Analyse der Struktur österreichischer F&E-Förderungen folgende Schlussfolgerungen für die Evaluierung von F&E-Maßnahmen in Österreich:

- Der österreichische Staat gibt im internationalen Vergleich relativ viel Geld – und das in zunehmendem Maße – für Subventionen aus. Das gilt insbesondere auch für Innovationssubventionen von Unternehmen. Kaum ein anderes Land gibt so viel Geld in % des BIP für die Förderung unternehmerischer F&E aus und finanziert gleichzeitig einen so hohen Anteil an den gesamten F&E-Ausgaben des Unternehmenssektors. Dieser hohe öffentliche Ressourceneinsatz erfordert und legitimiert auch den Ressourceneinsatz für die Evaluation der damit verbundenen Maßnahmen.
- Die Hälfte der Volumina der österreichischen F&E-Förderungen für Unternehmen wird in Form von steuerlichen Vergünstigungen von F&E-Aktivitäten (indirekte Forschungsförderung) vergeben. Diese in ihrer Bedeutung zunehmende Förderaktivität, wurde bislang nicht im Hinblick auf Effizienz und Effektivität überprüft.
- Während das Gesamtniveau an Innovationsförderungen für Unternehmen einen Spitzenwert aufweist, sind die österreichischen Ausgaben für die direkte Förderung von F&E von Unternehmen niedriger als in Frankreich, Spanien oder Schweden und nur geringfügig höher als in Finnland oder Norwegen.
- Aufgrund der hohen Förderintensität besteht in einigen Segmenten des Unternehmenssektors das Problem, dass ein Großteil der betreffenden Unternehmen eine Förderung erhält. Damit wird die Bildung einer Kontrollgruppe, die keine Förderung bezieht aber trotzdem innovativ ist, erschwert.

2.3. FFG-FÖRDERUNGEN UND THEMATISCHE ORIENTIERUNG DES BMVIT IM ÜBERBLICK

Die FFG vergibt ca. 60% der direkten österreichischen Forschungsförderung, gefolgt vom FWF mit etwa 30% und der AWS mit ca. 5% (Mayer et al. 2009). Bezieht man diese Werte nur auf die direkte Forschungsförderung für Unternehmen, so steigt die Bedeutung der FFG freilich weiter an.

Eine Darstellung der Aktivitäten der FFG hat die Pläne zur Restrukturierung der Förderstrategie (Themenmanagement) zu berücksichtigen. Wir stellen hier zunächst den nach wie vor aktuellen Status quo der programmorientierten Förderstrategie dar; auf die durch das Themenmanagement zu erwartenden Veränderungen wird in Kapitel 3.2.1 eingegangen.

Die FFG vergab 2010 Förderungen (inkl. Haftungen) in der Höhe von 554 Mio. €. Auf Akteursebene kam es damit zur Förderung von ca. 3000 Akteuren, wobei darunter Unternehmen ebenso wie Hochschulen fallen. Dieses Fördervolumen verteilt sich höchst ungleichmäßig auf ca. 40 Programme und mehr als 100 (!) Programmlinien (FTB 2011). Während also Österreich im internationalen Vergleich erst mit einiger Verzögerung das Steuerungsinstrument „Programm“ einsetzte, kam es im Anschluss daran zu einem beschleunigten Wachstum der Programmanzahl, was letztlich in eine stark fragmentierte Maßnahmenlandschaft unter dem Überbegriff der direkten Forschungsförderung mündete (Gottweis/ Latzer 2006). Diese Programmorientierung

löste die noch in den 1980er Jahren weithin dominante ressortfinanzierte Projektforschung ab (Biegelbauer 2010). Bedenkt man, dass auch die AWS eine Reihe von Programmen abwickelt, so repräsentiert die direkte Forschungsförderung in Österreich eine ausdifferenzierte, sich teilweise überschneidende Programmviefalt. Die Folge dieser Programmviefalt ist die Existenz von teilweise unterkritischen Programmgrößen, jedenfalls wenn man die oftmals sehr umfassenden Ziele in Betracht zieht, die diese Programme verfolgen. Darüber hinaus kam es durch die ständige Neueinführung von Programmen dazu, dass viele Programme erst eine relativ kurze Geschichte aufweisen. Diese Struktur der Programmviefalt in Österreich steht im deutlichen Gegensatz etwa zur Schweiz, die zwar einerseits wesentlich weniger Subventionen für F&E finanziert, aber andererseits auch eine klare Fokussierung auf einige wenige Programme aufweist (Arvanitis et al. 2010).

Damit ist die Evaluation von Programmen das bislang dominante Evaluationsdesign. Zusätzlich, aber wesentlich seltener finden umfangreiche Institutionenevaluierungen (z.B. FFF/FWF 2004) oder Systemevaluierungen (Aiginger et al. 2009) statt.

Diese strukturellen Kennzeichen der direkten, unternehmensbezogenen Forschungsförderung in Österreich tragen zunächst einmal dazu bei, dass von der Anzahl her viele Evaluierungen stattfinden – viele Programme, viele Evaluierungen. Die Frequenz wird weiterhin durch mehrmalige Evaluierungen im Lebenszyklus eines Programms erhöht (Ex-ante-, Interim- und Ex-post Evaluierung). Diese Konstellation birgt die Gefahr einer Evaluierungsmüdigkeit in sich.

Aus der Programmviefalt folgt weiterhin, dass Evaluierungen von ihrer Ressourcenausstattung her relativ „sparsam“ dimensioniert sind. So empfiehlt etwa die Plattform Forschungs- und Technologieevaluierung in ihren Standards, ein Evaluierungsbudget mit etwa 0,5-1% des Programmbudgets zu kalkulieren. Dies führt bei unterkritischen Programmgrößen dazu, dass meist einfache deskriptive Methoden angewandt werden. Die zahlreichen Programmneueinführungen der letzten Jahre erschweren die Wirkungsanalyse zusätzlich, weil längerfristige Entwicklungen oder wiederholte Wirkungsabschätzungen nicht möglich sind. Dies ist aber im FTI-Bereich aufgrund signifikanter time-lags zwischen Instrumenteneinsatz und Maßnahmeneffekt problematisch. Die Programmviefalt bedingt mögliche Probleme in der Zuordnung von Wirkungen auf einzelne Programme, wenn ein Unternehmen mehrere Förderungen gleichzeitig bezieht. Technisch gesprochen nimmt die Wahrscheinlichkeit von Messfehlern mit der Anzahl und Ähnlichkeit der Programme zu. Zuletzt ist darauf hinzuweisen, dass ein stark ausdifferenziertes Programmportfolio ceteris paribus zur Reduktion der Maßnahmenteilnehmer pro Programm beitragen. Dadurch werden quantitativ-analytische Methoden erschwert, weil diese möglichst hohe Fallzahlen erfordern.

Zusammen sorgen die Programmviefalt sowie die rasche Zunahme an Programmen damit für eine Reihe von Problemen bei der Anwendung innovativer Evaluationsmethoden. Das bereits erwähnte Themenmanagement könnte hier zu merklichen Verbesserungen beitragen.

Das BMVIT setzt im Rahmen seiner Förderstrategie eine Reihe von sektoralen Schwerpunkten. Nach Angaben der Homepage (<http://www.bmvit.gv.at/innovation/index.html> [27.10.2011]) umfassen diese folgende Schwerpunkte:

- Elektromobilität
- Energie- und Umwelttechnologie
- Informations- und Kommunikationstechnologien und Nanotechnologie
- Luftfahrttechnologie
- Raumfahrttechnologien
- Sicherheitsforschung (KIRAS)
- Verkehrstechnologie
- Produktion

Diese Schwerpunkte stehen im Einklang mit einem internationalen Mainstream, im Rahmen dessen sich die OECD-Staaten auf ähnliche Schwerpunkte festgelegt haben (OECD 2010). Generell hat die Frage nach der technologischen Ausrichtung eines Förderprogramms keine Folgen für die Evaluierung einschlägiger Maßnahmen. Relevant sind demgegenüber vor allem die Anzahl der Maßnahmenteilnehmer und die Anzahl der nichtteilnehmenden, aber vergleichbaren Unternehmen sowie die Programmziele. Ergibt sich etwa aufgrund von Branchenstrukturen, wie evtl. im Bereich der Luft- oder Raumfahrttechnologie, eine Situation, in der einerseits sehr wenige Unternehmen überhaupt in dieser Branche aktiv sind und andererseits eine Förderung an so gut wie alle diese Unternehmen erfolgt, so verunmöglicht dies etwa mikroökonomische Verfahren zur Abschätzung kausaler Fördereffekte.

3 FTI-Evaluierung in Österreich: Status quo und institutionelle Herausforderungen

„Evaluierung in der österreichischen Forschungs- und Technologiepolitik besitzt also eine stark konsensorientierte institutionelle Dimension, die aber im Wesentlichen auf dieses Politikfeld beschränkt bleibt, sehr stark von informellen Übereinkommen der Beteiligten untereinander gekennzeichnet ist und wenig Berührungspunkte zu Evaluierungsfunktionen anderer Politikfelder aufweist.“

Pichler (2009:47)

Kapitel 3 stellt Stand und Herausforderungen von Evaluationsaktivitäten FTI-politischer Maßnahmen in Österreich dar. Dabei wird neben den Interviewergebnissen auf weitere Datenquellen zurückgegriffen. Diese umfassen vor allem Policy-Dokumente, Strategiepapiere des Finanzministeriums sowie der FFG und quantitative Untersuchungen, welche den qualitativen Interviewergebnissen gegenübergestellt werden. Abschnitt 3.1 referiert die Position zentraler Stakeholder von FTI-Evaluierungen und Abschnitt 3.2 beschreibt und analysiert die mit der Haushaltsrechtsreform sowie mit der Einführung des Themenmanagements verbundenen Herausforderungen für FTI-Evaluierungen.

3.1. BESTANDS- UND BEDÜRFNISANALYSE ÖSTERREICHISCHER FTI-EVALUIERUNGEN

3.1.1 Viele und auch gute Evaluierungen

In Österreich wird viel und durchaus auch gut evaluiert – diese Aussage wurde in allen Interviews mehr oder weniger bestätigt. Insbesondere wurde die hohe Frequenz mit der Evaluierungen durchgeführt werden, hervorgehoben. Die Intensität der Evaluation österreichischer FTI-Politik kommt nicht zuletzt auch in der Systemevaluierung zum Ausdruck. Diese spezifische Form der Evaluierung wurde bislang erst in einigen wenigen Staaten in diesem Ausmaß durchgeführt (e.g. Finnland oder Tschechien).

Was die Qualität anlangt so zeigten sich auch hierbei alle befragten Akteure durchwegs zufrieden. Es wurde zwar eine zum Teil erhebliche Variabilität der Qualität von Evaluierungen erwähnt, aber letztlich stellen mangelhafte Evaluierungen die Ausnahme dar. Die Variabilität der Qualität trifft auch auf die Nachfrageseite zu. So wurde angemerkt, dass die unterschiedlichen Abteilungen in den FTI-relevanten Ressorts unterschiedliche Grade an Professionalität und Verbindlichkeit im Zusammenhang mit Evaluierungen aufweisen. Als ein Qualitätsindikator für die Nachfrageseite wurden mehrmals die Terms of References genannt. Zusammengenommen wurde aber in den Interviews betont, dass neben der generell zufriedenstellenden Qualität der Evaluierungen auch eine zunehmende Professionalisierung der Nachfrager und Anbieter von FTI-bezogenen Evaluierungen festgestellt werden kann. Dies äußert sich nicht zuletzt in einer verstärkten

Publikation von Evaluierungen, auch dann wenn diese kritische Ergebnisse liefern, wie dies etwa am Beispiel der Headquarter-Evaluierung exemplifiziert werden wird (Geyer/ Tiefenthaler 2011). Die Interviewergebnisse hinsichtlich der hohen Frequenz und generell zufriedenstellender Qualität von FTI-Evaluierungen sowie die zunehmende Professionalisierung der Stakeholder stimmen überein mit den Resultaten der 2006 durchgeführten Studie von Rhomberg et al. (2006), die ebenfalls eine Befragung von Anbietern und Nachfragern von FTI-Evaluierungen durchführten.

Auf die Frage nach der Qualität österreichischer FTI-Evaluierungen im internationalen Vergleich wurde von den meisten InterviewpartnerInnen zunächst festgehalten, dass dieser Vergleich schwierig sei, da der Wissensstand über die Lage in anderen Staaten aufgrund komplexer Informationslage tendenziell eher gering ist. Trotzdem wurde auch hier mehrheitlich festgestellt, dass Österreich im internationalen Vergleich gut abschneide und selbst ein good-practice-Modell darstellt. Insbesondere im Vergleich zur Situation in Deutschland wurden einige vorteilhafte Merkmale der österreichischen Evaluierungspraxis erwähnt. Angesichts dieser Antworten verwundert es nicht, dass die Frage nach Staaten, die als best-practice für Österreich dienen könnten, keine uneingeschränkt gültigen Beispiele von den InterviewpartnerInnen genannt wurden. Im Vergleich mit Staaten wie der Schweiz oder Großbritannien wurde auf die möglichen Gefahren einer eher mechanistischen, kennzahlen-basierten Evaluierungskultur hingewiesen.

3.1.2 Unbefriedigte Bedürfnisse und Verbesserungspotenzial

Bestehen in Österreich inhaltliche und/oder methodische Evaluierungsbedürfnisse, die bislang nicht durch das Angebot befriedigt werden konnten? Zunächst einmal ist festzustellen, dass das Mismatch zwischen den Bedürfnissen der Anbieter und jenen der Nachfrager ein Forschungsthema der Evaluationsforschung darstellt. Dabei stellt sich heraus, dass dieses Mismatch multidimensional ist. Darunter ist zu verstehen, dass zumindest vier Mismatches parallel existieren:

(1) PolitikerInnen wollen etwas anderes haben als EvaluatorInnen liefern können und/oder wollen. Dieser sogenannte „Delivery Gap“ ist in Tabelle 3 dargestellt. Von Seiten des BMVIT wurde aber keiner der hier genannten Aspekte als relevant für die österreichische Situation klassifiziert. Es kann generell festgestellt werden, dass die interviewten Akteure der FTI-Politik durchaus realistische Vorstellungen und Erwartungen z.B. hinsichtlich der Schwierigkeit und Problematik von kausalen Wirkungsabschätzungen („clear attribution of effects to investment“) und kennzahlenbasierten Evaluationsmethoden haben. Freilich müssen diese Ergebnisse keineswegs repräsentativ sein, da eine positive Selektion der InterviewpartnerInnen anzunehmen ist.

Tabelle 3: Der "Delivery Gap"

What policymakers want	What evaluators say
<ul style="list-style-type: none"> • Information in time on spending decision • Clear attribution of effects to investment • Independent evidence of research excellence • Key indicators to monitor and benchmark 	<ul style="list-style-type: none"> • Research may take years to have effects • Linear model is a rare case and additionally is complex to assess • Peers defend their subject field and international colleagues • Crude regime distorts performance and can be manipulated

Zinöcker 2007

(2) Das Mismatch zwischen Politik und Evaluationsanbietern existiert auch in umgekehrter Richtung, wobei hier andere Dimensionen Gegenstand des Mismatch sind bzw. sein können. Tabelle 4 stellt diese Dimension dar. Die Interviews mit EvaluierungsanbieterInnen bestätigten vor allem den letzten und vorletzten der aufgelisteten Punkte. Bezüglich Ressourcen wurde angemerkt, dass die Anforderungen, wie sie in den Terms of References (TOR) definiert sind, bisweilen in keiner vernünftigen Relation zur Auftragssumme stehen.³ Der vierte Punkt ist ebenso Gegenstand des Mismatch zwischen den Erwartungen von ProgrammmanagerInnen und EvaluatorInnen.

Tabelle 4: Der "Customer Gap"

What evaluators want	What policy makers say
<ul style="list-style-type: none"> • Clearly defined and hierarchical objectives • Guaranteed independence • Time and resources to do the job • Full access to information and stakeholders 	<ul style="list-style-type: none"> • Programmes are a compromise involving multiple and conflicting objectives • Recommendations must be within realistic policy constraints • We need the results in three months • Everyone is overworked and busy

Zinöcker 2007

(3) Zwischen den ProgrammmanagerInnen und den EvaluatorInnen gibt es ebenfalls unterschiedliche Ansprüche, die insbesondere auch das Thema Datenverfügbarkeit und Datenqualität (Tabelle 5).

Diese Problematik stellt tatsächlich einen kritischen Punkt für die methodische Weiterentwicklung österreichischer FTI-Evaluierung dar. Kapitel 5 diskutiert diesen Aspekt ausführlich, auch im internationalen Vergleich.

³ So merkte etwa ein InterviewpartnerInnen an, dass es eben unmöglich sei, „die halbe Welt um 40.000 € zu erklären“.

Tabelle 5: Der "Management Gap"

What program managers say	What evaluators answer
<ul style="list-style-type: none"> • I had tons of work and a lot of customer relationships • Look at this nice development • We collected lots of facts about our projects 	<ul style="list-style-type: none"> • Fine, where is the documentation? • What are the social returns? • Not a single number is a useful additionally measure!

Zinöcker 2007

(4) Die Interviewergebnisse haben ein weiteres, viertes Mismatch gezeigt, welches sich bislang nicht in der einschlägigen Literatur wiederfindet und möglicherweise auch ein Spezifikum der österreichischen institutionellen Arbeitsteilung im FTI-Bereich ist. Es handelt sich hierbei um die unterschiedlichen Ansprüche und Erwartungen an Evaluierungen zwischen ProgrammmanagerInnen bzw. Agenturen und Ministerien.

Im Wesentlichen besteht dieses Mismatch darin, dass das Ministerium als beauftragende und damit zahlende Institution laut Aussagen einiger InterviewpartnerInnen, die evaluatorischen Bedürfnisse der FFG zu wenig berücksichtigt. Es wurde erwähnt, dass eine Einbeziehung der FFG-Programmmanger in die Verfassung der Terms of References erwünscht, aber nicht stets realisiert ist. Demgegenüber wurde von Seiten des BMVIT erklärt, dass die FFG eigene Evaluierungen, die sich explizit den für sie interessanten Fragestellungen widmen, in Auftrag geben könne. Tatsächlich hat die FFG in den letzten Jahren zunehmend weniger Evaluierungen in Auftrag gegeben.

Weitere in den Interviews genannten unbefriedigten Evaluierungsbedürfnisse bzw. Bereiche mit Verbesserungspotenzial umfassen folgende Aspekte:⁴

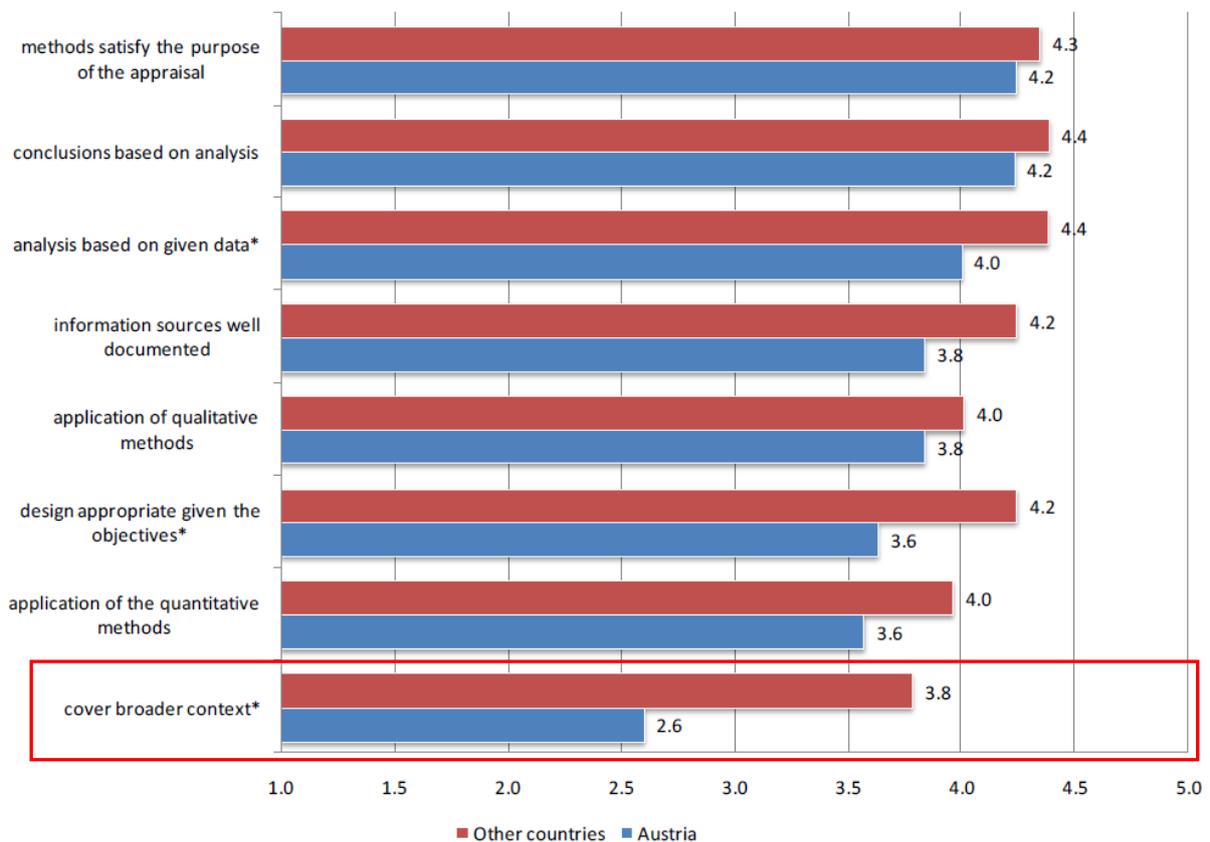
- Mangel an ex-post Evaluierungen. Es gibt sehr wenig Studien oder Evaluierungen, die sich langfristig Gedanken machen, welche Effekte dieses Programm gehabt hat. Diese Einschätzung wird durch die Ergebnisse der Studie INNO-APPRAISAL bestätigt⁵: Die Mehrzahl der Evaluierungen hat formativen Charakter und wird in den frühen Phasen des Lebenszyklus eines Programmes eingesetzt. Selten werden summative Evaluierungen durchgeführt.
- Mangel an Wirkungsanalysen. In den Interviews wurde ein Mangel im Sinne der Beantwortung kausaler Programmeffekte angemerkt.
- Mangel an evaluatorischer Information zu konkreten Fragen des Fördermanagements. Einige InterviewpartnerInnen stellten fest, dass Evaluationen tendenziell zu abstrakt bzw. auf einer zu hohen Aggregationsebene ansetzen. Damit verbunden ist die Nichtbeachtung der konkreten Arbeit der Fördergeber („Förderhandwerk“), die jedoch auch evaluiert werden soll.

⁴ Diese Aspekte wurden von den verschiedenen Gruppen von befragten Stakeholdern genannt. Für eine weitere Beschreibung und Dokumentation der Ergebnisse siehe <http://www.proinno-europe.eu/appraisal> [03.11.2011]

⁵ Das Ziel dieser 2010 publizierten Studie war eine Bestandsaufnahme der FTI-Evaluierungspraxis in den EU-Staaten.

- Mangel an Portfolioevaluierung. Eine Mehrzahl der InterviewpartnerInnen sprach sich für eine verstärkte Berücksichtigung folgender Fragestellungen aus: Wie ist das zu evaluierende Programm relativ zu den anderen am Fördermarkt positioniert? Ist der Politik-Mix zur Adressierung einer spezifischen innovationspolitischen Zielsetzung adäquat? Diese Einschätzung deckt sich mit den Ergebnissen von INNO-APPRAISAL (2010). Wie Abbildung 7 zeigt, ist der für Österreich negative Unterschied bei der Qualitätseinschätzung von Evaluierungen durch FTI-politische Akteure bei dem Merkmal „Erfassung des weiteren Programmumfelds“ besonders groß, relativ zu den Vergleichsländern der EU. Auch im Vergleich zu allen anderen Qualitätsmerkmalen zeigt sich, dass dieses Merkmal mit Abstand die geringste Zufriedenheit aufweist.

Abbildung 7: Qualität der Evaluationen



INNO-APPRAISAL 2010

- Mangel an „guten“ Fallstudien. In den Interviews wurde einerseits die Schwierigkeit quantitativer Kausalanalyse betont, aber andererseits die Möglichkeit erwähnt, mittels repräsentativer Fallstudien zumindest „kausale Erzählungen“ zu produzieren. Dies sei mittels eingehender Beschäftigung mit dem Fallstudienobjekt grundsätzlich möglich. Als positives Beispiel wurde die Headquarter-Evaluierung erwähnt (Geyer/ Tiefenthaler 2011).
- Mangel an Effizienzanalysen. Seitens des BMF wurde ein deutlicher Mangel an Effizienzbetrachtungen in FTI-Evaluierungen argumentiert. Es sollten demnach vermehrt auch die Kosten der Abwicklung eines Programms analysiert und mittels Vergleich hinsichtlich ihrer Effizienz beurteilt werden. Auch dieses Ergebnis stimmt gut mit den Resultaten von INNO-

APPRAISAL (2010) überein: Tatsächlich ist Effizienz der am wenigsten adressierte Themenbereich in österreichischen FTI-Evaluationen und der Abstand zu den Vergleichsländern ist sehr groß. Allerdings zeigt sich auch im internationalen Vergleich, dass Effizienzbetrachtungen von untergeordneter Bedeutung sind.

- Mangel an Verbindlichkeit. Nach Aussagen von InterviewpartnerInnen besteht weiterhin Verbesserungsbedarf bei der Verwertung von Evaluierungsergebnissen. Generell besteht hier ein großer diskretionärer Spielraum seitens der beauftragenden Abteilung im jeweiligen Ministerium, der entsprechend unterschiedliche Umgangsformen mit Evaluierungsergebnissen nach sich zieht. Dieser Mangel an Verbindlichkeit setzt sich fort in der Frage nach der Publikation von Evaluierungen. Wenngleich hier in Österreich mit der fteval eine einzigartige Plattform mit dem Ziel der Publikation aller Evaluierungen etabliert werden konnte, werden nach wie vor nicht alle Evaluierungen veröffentlicht. Ein internationaler Vergleich zeigt, dass etwa die finnische Agentur Tekes oder die belgische Agentur iwt eine Politik der uneingeschränkten Publikation von Evaluationsergebnissen folgen.

Die zu Beginn der Studie formulierte Hypothese eines Mismatch zwischen Instrumentencharakteristika (e.g. Kooperationsförderung vs. klassisches Einzelprojekt) und Evaluierungsmethoden wurde von keinem InterviewpartnerInnen als zutreffend bestätigt.

Im Gegensatz zu den Ergebnissen der Befragung von 2006 (Rhombert et al. 2006), die ein Bedürfnis nach vermehrter makroökonomischer Analyse (z.B. Auswirkungen der FTI-Politik auf Beschäftigung und Produktivität) identifizierten, wurde eine unbefriedigte Nachfrage nach makroökonomischer Analyse bei den 2011 durchgeführten Interviews *nicht* geäußert. Tatsächlich konnte auch durch weitere Recherche keine einzige FTI-Evaluierung in Österreich ausfindig gemacht werden, welche die Auswirkungen von FTI-Politik auf makroökonomische Größen evaluiert hat.

3.2. INSTITUTIONELLE HERAUSFORDERUNGEN, EVALUIERUNGSDESIGN UND ZIELKONFLIKTE

3.2.1 Themenmanagement

Das Themenmanagement kann als eine organisatorische Innovation im Fördermanagement von FFG und BMVIT verstanden werden, die gegenwärtig noch in Planung ist. Es handelt sich dabei u.a. um eine Reaktion auf die Ergebnisse der Systemevaluierung. Folgende Problemstellungen sollen damit adressiert werden: Erstens hat die Programmvielelt mittlerweile ein Niveau erreicht, das die Frage nach der Steuerbarkeit des hochgradig ausdifferenzierten Programmportfolios aufwirft. Zweitens hat die Differenzierung der Programme nicht zur klaren Abgrenzung der Zielgruppen geführt. Für Unternehmen sind viele Programme vielmehr nahe Substitute dar. Gegeben diese differenzierten Anreize versuchen Unternehmen die maximalen Förderressourcen einzuwerben, womit de facto Förderprogramme in eine Form der Konkurrenz zueinander gesetzt werden. Diese Konstellation trifft insbesondere auch auf das Nebeneinander von thematischen Förderprogrammen und Basisprogrammen zu. Drittens führte die Programmvielelt auch zu sehr

heterogenen, inhaltlich schwer begründbaren Governance-Arrangements. Während einige Programme relativ eng von den Ministerien abhängen, ist dies bei anderen Programmen nicht der Fall. Als Resultat dieser Befundlage soll mit dem Themenmanagement, ganz allgemein gesprochen, eine Vereinheitlichung und Vereinfachung erzielt werden. Gleiches soll gleich behandelt werden, insbesondere hinsichtlich Ausschreibungsmodalitäten, Förderkonditionen und Bewertungsverfahren. Weiterhin soll durch einen koordinierten Ausschreibungskalender die Planungssicherheit auf Seiten der Fördernehmer erhöht werden.

Anstatt eines Programmmanagements vieler kleiner Programme soll das Themenmanagement auf Basis eines standardisierten Instrumentenkoffers einen jeweils passenden Mix für unterschiedliche thematische Vorgaben und Zielsetzungen des BMVIT zusammenstellen. Tabelle 6 zeigt einen Entwurf für den Instrumentenkoffer. Demnach werden 6 Gruppen von Instrumenten unterschieden. Es gibt nunmehr 16 unterschiedliche Instrumente, die u.a. nach Größe differenziert werden.

Berücksichtigt man alle weiteren Differenzierungen ergibt sich in Summe eine Anzahl von 25 Instrumenten. Tabelle 7 zeigt eine Zuordnung von Programmen zu den Instrumenten. Demnach werden 41 Programme 16 bzw. 25 Instrumenten zugeordnet.

Die Auswirkungen eines Themenmanagements auf die Evaluierungspraxis wären beträchtlich. Bedenkt man, dass die bisherigen Evaluierungen zu einem Großteil Programmevaluierungen sind, so ist eine Abkehr von der Programmlogik eine bedeutsame Veränderung. Diese Einschätzung wurde von allen InterviewpartnerInnen bestätigt. Grundsätzlich lassen sich Auswirkungen auf die unterschiedlichen Zeitpunkte der Evaluation unterscheiden. Die Idee einer Themensetzung durch das BMVIT könnte etwa die Nachfrage nach ex-ante Evaluierungen erhöhen. Zumindest theoretisch sind damit eine ganze Reihe von Problemen verbunden, die mit dem Begriff „targeting“ umschrieben werden können (siehe hierzu Gassler/ Polt 2006). Um dieses Problem zu vermeiden könnten zwar die Themen sehr breit definiert werden, andererseits könnte damit wiederum der Hebel öffentlicher Subventionen zu klein werden, um substantielle Impulse setzen zu können. Besonders kritisch ist der Zusammenhang zwischen Themen und Branchen. Es besteht die Gefahr, dass bestimmte Branchen sich einzelne Themen in monopolistischer Art und Weise aneignen (z.B. Umwelttechnologien wird durch die Photovoltaik-Industrie vereinnahmt) und so damit zusammenhängenden politökonomischen Probleme wie etwa „regulatory capturing“ oder „rent seeking“ auftreten. Letzteres ist nicht zuletzt auch eine Funktion der Förderdauer eine weitere zu entscheidende Frage bei der Implementierung eines neuen Themas.

Tabelle 6: Instrumentenkoffer Entwurf 1: Förderinstrumente und Förderkonditionen (Stand August 2011)

Förderinstrumente						
Gruppe	Instrument		Förderintensität	Fördervolumen	Verfahren	Rechtsgrundlage EU/national
Einstieg	C1	Projektvorbereitung	50%	≤ 3 k EURO	M1	Gemeinschaftsrahmen / FFG-RL
		XS Innovationscheck/klein	100%	5 k EURO	M1	De-minimis-VO / SonderRL
	C2	S Innovationscheck/groß	80%	≤ 10 k EURO	M1	De-minimis-VO / SonderRL
		M Feasibility	40 – 75%	≤ 30 k EURO	M2	Gemeinschaftsrahmen / FFG-RL
		L Sondierung	40 – 80%	≤ 200 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
FEI-Projekt	C3	E¹ Einzelprojekt - Entwicklung	25 – 60%	≤ 2.990 k EURO	M2	Gemeinschaftsrahmen / FFG-RL
		I² Einzelprojekt - Forschung	45 – 70%	≤ 2.000 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
		G³ Einzelprojekt - Forschung/ESA	100%	≤ 2.000 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
	C4	E – I Kooperationsprojekt	35 – 80%	≤ 2.000 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
		G Kooperationsprojekt – ESA	100%	≤ 2.000 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
	C5 Leitprojekt		40 – 80%	>2.000 k EURO	M4	Gemeinschaftsrahmen / FTE-RL
	C6 Wissenschaftstransfer		60 – 75%	≤ 750 k EURO	M3	Gemeinschaftsrahmen / FFG-RL
Markteinführung	C7 Markteinführungsprojekt		10 – 50%	≤ 500 k EURO	M3	AGVO / SonderRL
Struktur	C8 Zentrum		40 – 80%	nach Zentrumsart	M4	Gemeinschaftsrahmen / FTE-RL oder SonderRL
	C9 Strukturaufbau		40 – 70%	≤ 1.500 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
	C10 Innovationsnetzwerk		40 – 75%	≤ 750 k EURO	M3	Gemeinschaftsrahmen / FTE-RL
	C11 Organisation & Personalentw.		50 – 70%	≤ 50 k EURO	M1	De-minimis-VO / SonderRL
Personen	C12	S Praktikum/SchülerInnen	50 – 100%	≤ 1 k EURO/Prakt.	M1	AGVO / SonderRL
		M Praktikum/StudentInnen	50%	≤ 11 k EURO/Prakt.	M1	AGVO / SonderRL
		L Dissertationen	50%	≤ 100 k EURO/Diss.	M1	AGVO / SonderRL
	C13 Grants		80 – 100%	≤ 5 k EURO	M1	- / SonderRL
	C14 Netzwerk-Forschung-Schule		100%	≤ 60 k EURO	M3	De-minimis-VO / SonderRL
	C15	S Qualifizierungsnetzwerk/klein	60 – 100%	≤ 50 k EURO	M1	AGVO / SonderRL
		M Qualifizierungsnetzwerk/mittel	60 – 100%	≤ 500 k EURO	M3	AGVO / SonderRL
L Qualifizierungsnetzwerk/groß		60 – 100%	≤ 1.000 k EURO	M4	AGVO / SonderRL	
EU-Kofinanzierung	C16 EU-Kofinanzierung		75%	offen	M1	AGVO / SonderRL

Finanzierungsinstrument					
Aufwendungen	C17 F&E-Dienstleistungen		100% Finanzierung	M3	§ 10 Z13 BVergG 2006

¹ experimentelle Entwicklung ² industrielle Forschung ³ Grundlagenforschung (FFG)

Tabelle 7: Instrumentenkoffer Entwurf 2: Instrumente und zugeordnete Programme (Stand August 2011)

Förderinstrumente			
Gruppe	Instrument		Aktuelles Angebot/Formate
Einstieg	C1	Projektvorbereitung	1. Projektstart, Projektstart EU (BP)
		XS Innovationssscheck/klein	2. Innovationssscheck
	C2	S Innovationssscheck/groß	3. Innovationssscheck Plus
		M Feasibility	4. Feasibility (BP)
		L Sondierung	5. Technische Durchführbarkeitsstudien (TP)
FEI-Projekt	C3	E¹ Einzelprojekt - Entwicklung	6. Einzelprojekt (BP) 7. Headquarters 8. Young Experts 9. EUREKA-Projekt
		I² Einzelprojekt - Forschung	10. Einzelprojekt (Energien 2020) 11. FEMtech Forschungsprojekte (Talente)
		G³ Einzelprojekt – Forschung/ESA	12. Einzelprojekt ASAP
	C4	E – I⁴ Kooperationsprojekt	13. Kooperationsprojekte (TP) 14. EUREKA-Projekt, ERA-Net Projekte 15. Kooperationsprojekt ASAP 16. FEMtech Forschungsprojekte (Talente)
		G Kooperationsprojekt – ESA	17. Kooperationsprojekt ASAP 18. GE-NAU-Kooperationsprojekte
	C5	Leitprojekt	19. Leitprojekte bzw. Leuchtturmprojekte (TP) 20. Verbund (Nanoinitiative)
	C6	Wissenschaftstransfer	21. Brückenschlag 1 / BRIDGE
	Markteinführung	C7	Markteinführungsprojekt
Struktur	C8	Zentrum	23. COMET: K2, K1 24. Laura Bassi Centres of Expertise 25. AplusB
	C9	Strukturaufbau	26. RSA – Research Studios Austria 27. Josef Ressel Zentren 28. COMET: K-Projekte 29. COIN – Aufbau
	C10	Innovationsnetzwerk	30. COIN – Kooperation und Netzwerke
	C11	Organisation & Personalentw.	31. FEMtech Karriere (Talente)
	Personen	C12	S Praktikum/SchülerInnen
M Praktikum/Studentinnen			33. FEMtech Praktika für Studentinnen (Talente)
L Dissertationen			34. Praktika für DissertantInnen (Talente) 35. Stipendien in thematischen Programmen
C13		Grants	36. Karriere-Grants (Talente)
C14		Netzwerk-Forschung-Schule	37. Talente regional (Talente)
C15		S Qualifizierungsnetzwerk	38. Qualifizierungsseminare (FOKO)
		M Qualifizierungsnetzwerk	39. Qualifizierungsnetze (FOKO)
	L Qualifizierungsnetzwerk	40. Innovationslehrveranstaltungen mit tertiärem Charakter (FOKO)	
EU-Kofinanzierung	C16	EU-Kofinanzierung	41. TOP.EU
Finanzierungsinstrument			
Aufwendungen	C17 F&E-Dienstleistungen	Forschungsaufträge thematische Programme	

Wie auch immer die Themenauswahl letztlich stattfindet, sie wird sich diesen Problemen stellen müssen. Rationale Politik wird sich dabei wissenschaftlich in Form von ex-ante Evaluierungen beraten lassen.

Neben der Themenauswahl stellt sich noch die Aufgabe der Wahl der Instrumente: Welches Instrumentenportfolio soll für ein spezifisches Thema ausgewählt werden? Erneut wäre hier eine Prognose über die Struktur der potentiellen Fördernehmer zu treffen, auf deren Basis Instrumente auszuwählen sind. Besondere Herausforderung stellt in diesem Zusammenhang sicherlich die Einschätzung von Interaktionseffekten zwischen den Instrumenten dar. Nach Vonortas et al. (2007: 679) gehört diese Fragestellung zu den aktuellen Herausforderungen der Evaluationsforschung im FTI-Bereich: „Policy makers are increasingly aware that the success of their efforts to finance and promote long-term research is dependent, not only on individual programmes, institutions and infrastructure, but also on ‘portfolios’ and ‘systems’ of programmes and interventions that interact.“ Diese Frage ist nicht zuletzt auch deshalb von Bedeutung, weil Effekte im Sinne einer „behavioral additionality“ nicht durch ein einziges Instrument, sondern nur durch die Kombination und den wiederholten Einsatz unterschiedlicher Instrumente erzielt werden können: „When one acknowledges that the success of policy intervention does not solely depend on future funding schemes, it becomes clear that future work should focus more on the evaluation of programme portfolios and their interactions.“ (Falk 2006:676). Freilich ist die Frage nach Interaktionseffekten der Förderung auch für interim- und ex-post Evaluierungen von Relevanz.

Durch das Themenmanagement kommt es möglicherweise zum Ende der klassischen Programmevaluierung. Aus der neuen Konstellation lassen sich zwei grundsätzliche Evaluierungsstrategien ableiten:

- Instrumentenevaluierung: Ein Instrument wird über mehrere Themen hinweg evaluiert
- Themenevaluierung: Mehrere Instrumente werden in Bezug auf ein Thema evaluiert

Zieht man die Interventionslogik als Maßstab zur Selektion des Evaluationsdesigns heran, so erscheint die Themenevaluierung als adäquate Evaluationsperspektive. Auf der anderen Seite legt die Standardisierung der Instrumente sowie das Ziel, durch den gemeinsamen Einsatz von Instrumenten Synergieeffekte der Förderung zu erzielen, auch eine Instrumentenperspektive nahe. Freilich sind diese beiden Zugänge nicht im Sinne eines entweder/oder zu verstehen, sondern als komplementär.

Bei der Instrumentenevaluierung entsteht allerdings ein Problem, welches als Heterogenitätsproblem bezeichnet werden könnte. Gemäß Tabelle 7 werden einzelnen Instrumenten bis zu vier Programme zugeordnet. So basieren z.B. die Programme bzw. Formate Einzelprojekt (BP), Headquarters, Young Experts und EUREKA-Projekt alle auf dem gleichen Instrument „Einzelprojekt-Entwicklung“. Eine ähnliche Heterogenität zeigt das Instrument „Zentrum“. Hier werden die K-Programme, das AplusB Programm und die Laura Bassi Centres of Excellence zusammengefasst. Diese Programme lassen sich nicht einfach einheitlich evaluieren, nur weil sie das gleiche Instrument nutzen.

So sind etwa mögliche Erfolgsfaktoren für AplusB Zentren sehr verschieden von jenen der Kompetenzzentren. Auch sind die Erfolgsgrößen keineswegs identisch.

Somit kann festgestellt werden, dass eine Instrumentenevaluierung im engeren Sinne eine hinreichende Homogenität der Programme bzw. Formate voraussetzt, welche das entsprechende Instrument anwenden.

Ähnliche Probleme lassen sich auch für Themenevaluierungen identifizieren, wenngleich die Einheitlichkeit durch das Thema sowie eine übergeordnete Zielformulierung ein Stück weit realisiert werden könnte. Klar ist aber auch, dass eine Themenförderung sehr unterschiedliche Fördernehmer ansprechen wird: Unternehmen, Universitäten, ArbeitnehmerInnen, Zivilgesellschaft, etc. Für die unterschiedlichen Fördernehmer gelten wiederum unterschiedliche Erfolgsgrößen. Weiterhin wird die Möglichkeit einer Kontrollgruppenbildung vielleicht noch stärker eingeschränkt, wenn etwa alle relevanten Akteure eines Themas gefördert werden. Die Zahl der Ziele wird tendenziell zunehmen, wenngleich nicht proportional, und vermehrt auch auf sozioökonomische Ziele Rücksicht genommen werden.

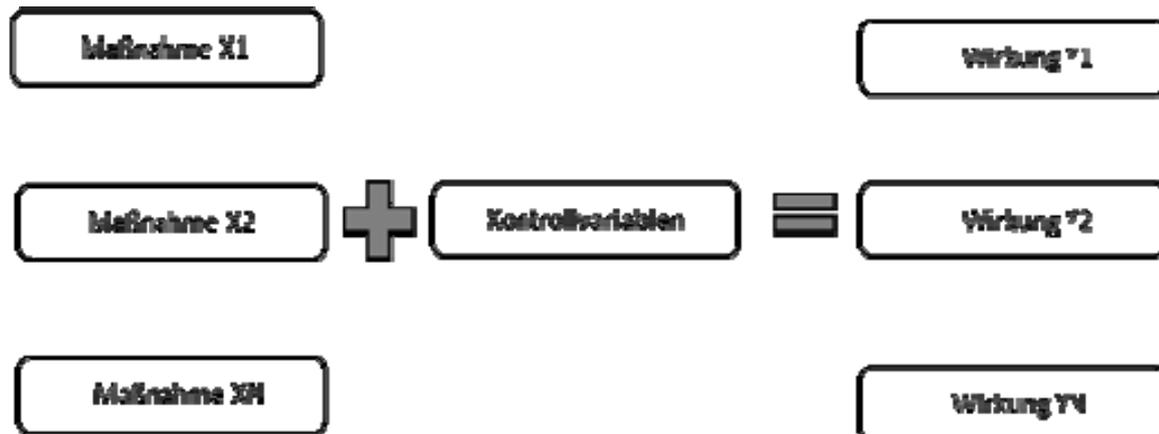
Zusammenfassend kann festgestellt werden, dass die Komplexität von Evaluierungen durch die Einführung des Themenmanagements tendenziell zunehmen wird (siehe Abbildung 8 und Abbildung 9).

Abbildung 8: Klassische Programmevaluierung



Eigene Darstellung

Abbildung 9: Steigende Komplexität durch Portfolio- bzw. Themenevaluierung



Eigene Darstellung

Diese Zunahme an Komplexität könnte eine weitere Verschiebung in Richtung qualitativer Methoden fördern. Ökonometrische Analysen werden daher möglicherweise noch weiter an Bedeutung verlieren.

Das klar abgegrenzte einzelne Förderprogramm das Standardbeispiel zur Anwendung ökonomischer Methoden dar. Wenn schon bei bisherigen Programmevaluierungen so gut wie keine methodisch anspruchsvolleren, quantitativen Methoden zum Einsatz kamen, so ist damit auch nicht bei der anspruchsvolleren Themenevaluierung zu rechnen.

Durch Aggregation aller Förderungen in monetäre Größen könnte eine ökonomische Abschätzung kausaler Effekte möglich sein. Ohne Kontrollgruppenbildung ist jedenfalls ein metrisches Skalenniveau der Fördervariablen notwendig. Dieser Ansatz würde der Modellierung in Streicher (2007) entsprechen. Es stellt sich jedoch die Frage, ob die Anzahl der geförderten Unternehmen pro Thema groß genug ist, um eine derartige Analyse zu ermöglichen. Auf die Problematik einer möglichen Kontrollgruppenbildung wurde bereits hingewiesen. Als illustratives Beispiel für die Problematik der Förderaggregation mag ein Unternehmen gelten, das einerseits Förderungen in Form Einzelprojektförderung erhält und gleichzeitig an einem Kompetenzzentrum beteiligt ist. Während Einzelprojektförderung eine klare Angabe über die monetäre Förderhöhe ermöglicht, ist dies bei Kompetenzzentren generell nicht möglich. Hier kann keine einfache Zurechnung monetärer Förderungen auf einzelne Unternehmen erfolgen.

Eine andere Möglichkeit zur quantitativen Analyse würde sich in Form von branchenspezifischen Produktionsfunktionen anbieten. Hier liegen entsprechende Sekundärdaten jedenfalls prinzipiell vor. Allerdings wirft dies wieder die Frage nach dem Zusammenhang von Branchen und Themen auf. Ein mögliches Resultat der durch die Themenevaluierung zu erwartenden methodischen Verschiebungen könnte ein Zielkonflikt zwischen „harter“ Outputorientierung und „weicher“ Themenförderung sein. Ersteres lässt sich zumindest auf Einzelprogrammebene ansatzweise begründen, während die Vielzahl an Instrumenten, Zielen und Akteuren eine derartige Perspektive per se eher ausschließt.

Während die konkrete Ausgestaltung von Evaluierungsdesigns offen bleiben muss, ist die unmittelbare Folge zunächst eine höchstwahrscheinlich geringere Anzahl an Evaluierungen bei gleichzeitiger Zunahme der finanziellen Mittel je Evaluierung. Dies ist positiv zu beurteilen, da damit erstens die Fixkosten pro Evaluierung relativ abnehmen und mehr Ressourcen in kreativere Evaluierungsansätze gelenkt werden können, zweitens wirkt dies auch einer latenten Evaluierungsmüdigkeit entgegen, die durch die zahlreichen Programmevaluierungen mit verursacht wird. Drittens wird die Forderung nach einer Portfolioevaluierung ein Stück weit realisiert.

3.2.2 Wirkungsorientierte Haushaltsführung

Die Einführung einer Wirkungsorientierten Haushaltsführung (WHF) ist Bestandteil der zweiten Etappe der Haushaltsrechtsreform, die mit 1.1.2013 in Kraft treten wird. Die erste Etappe wurde bereits 2009 umgesetzt. Beide Teile der Haushaltsrechtsreform wurden am 06.12.2007 vom Nationalrat einstimmig beschlossen. Die theoretische Legitimation dieser Reformen beruht auf Ideen des „New Public Management“ (NPM). NPM ist ein umfassender Ansatz zur Reorganisation der Art und Weise der Leistungserfüllung und Planung des öffentlichen Sektors. Insbesondere geht es um eine Abkehr vom „klassisch bürokratischen Regierungsmodell“ (Nowotny, Zagler 2009).

Ein wesentliches Unterscheidungsmerkmal ist die Priorisierung einer output-orientierten anstelle einer bislang dominanten input-orientierten Steuerung („Performance Budgeting“ statt Kameralistik) (Nowotny, Zagler 2009). Die österreichische Haushaltsrechtsreform reiht sich somit in eine Reihe von Reformen in anderen Staaten ein, die bereits ein Haushaltswesen nach den Prinzipien des NPM eingeführt haben, z.B. die USA oder die Schweiz.

Ein für die österreichische FTI-Politik interessantes Beispiel in diesem Zusammenhang sind die Debatten rund um die Einführung des Government Performance and Results Act in den USA. Einige der in Box 3 skizzierten Problemlagen in den USA wurden auch in den Interviews zu dieser Studie geäußert. Aktuell wird unter dem Titel „Star Metrics“ gerade wieder eine umfangreiche Reform zur Erfassung der Effekte öffentlicher F&E-Ausgaben in den USA diskutiert (siehe dazu Lane 2009 und 2010, Macilwain 2009).

Box 3: Fallbeispiel Government Performance and Results Act (GPRA), USA, 1993

In 1993, the U.S. Congress quietly passed a good government bill, with little fanfare and full bipartisan support. President Clinton happily signed it into law (...). The law required each part of government to set goals and measure progress toward them as part of the process of determining budgets. “The first reaction of the federally supported scientific research community to this law was shock and disbelief. (...) The law required strategic planning and quantitative annual performance targets--activities either unknown or unwanted among many researchers.

The discussion in the research community quickly converged on a set of inherent problems in applying the GPRA requirements to research. First, the most important outcomes of research, major breakthroughs that radically change knowledge and practice, are unpredictable in both direction and timing. Trying to plan them and set annual milestones is not only futile but possibly dangerous if it focuses the attention of researchers on the short term rather than the innovative. Second, the outputs of research supported by one agency intermingle with those of activities supported from many other sources to produce outcomes. Trying to line up spending and personnel figures in one agency with the outcomes of such intermingled processes does not make sense. Third, there are no quantitative measures of research quality. (...)

The risks of applying GPRA too mechanistically in research thus became clear. First, as we have seen, short-termism lurks around every corner in the GPRA world in the form of overemphasis on management processes, on research outputs ("conduct five intensive operations periods at this facility") rather than outcomes ("improve approaches for preventing or delaying the onset or the progression of diseases and disabilities"), and on the predictable instead of the revolutionary. Short-termism is probably bad in any area of government operations but would be particularly damaging in research, which is an investment in future capabilities.

Cozzens (1999)

Für die FTI-Politik bzw. die Evaluierung derselben sind dabei neben der WHF auch die Aspekte betreffend Geschlechtergerechtigkeit („Gender Budgeting“), die Teil der 2. Etappe der Haushaltsrechtsreform sind, von Bedeutung.

Es geht demnach um die „Verankerung des Grundsatzes der Wirkungsorientierung unter Berücksichtigung der tatsächlichen [sic!] Gleichstellung von Frauen und Männern im gesamten Kreislauf der Haushaltsführung“.⁶ Box 4 fasst die zentralen Ziele der zweiten Etappe der Haushaltsrechtsreform zusammen.

Box 4: Ziele der zweiten Etappe der Haushaltsrechtsreform

- Transparente Darstellung gegenüber Parlament und Öffentlichkeit, welche Wirkungsziele angestrebt werden und welche Maßnahmen mit den zur Verfügung stehenden Ressourcen gesetzt werden, um diese zu erreichen
- Erleichterung der Prioritätensetzung auf Ebene der Politik und davon abgeleitet in der Verwaltung
- Förderung einer stärkeren Ergebnisverantwortlichkeit der Ressorts und der einzelnen haushaltsführenden Stellen
- Ein zentrales Ziel der zweiten Etappe der Haushaltsrechtsreform ist die Verknüpfung der Kosten der Verwaltung mit den von ihr erbrachten Leistungen anstatt der bloßen Orientierung an den budgetierten Mitteln.
- Ein wesentlicher Aspekt der Wirkungsorientierung ist auch das "Gender Budgeting", das mit der Haushaltsrechtsreform in Österreich eingeführt wird und dessen gesetzliche Grundlage bereits in der Bundesverfassung verankert ist. Bund, Länder und Gemeinden werden verpflichtet, bei der Haushaltsführung die tatsächliche Gleichstellung von Frauen und Männern anzustreben.

http://m.bmf.gv.at/Budget/Haushaltsrechtsreform/2EtappederHaushalts_10081/_start.htm [16.11.2011].

Auffallend an den Zielen der zweiten Etappe der Haushaltsrechtsreform ist die Betonung von Effizienzaspekten. Es geht demnach nicht nur um Orientierung am Output, sondern auch um eine Berücksichtigung des Input, der zur Erreichung des Outputs notwendig sind. Eine Überprüfung dieser Effizienzziele würde eine Effizienzanalyse erfordern, Effizienzanalysen werden jedoch in der Evaluation von F&E-Maßnahmen eher selten durchgeführt.

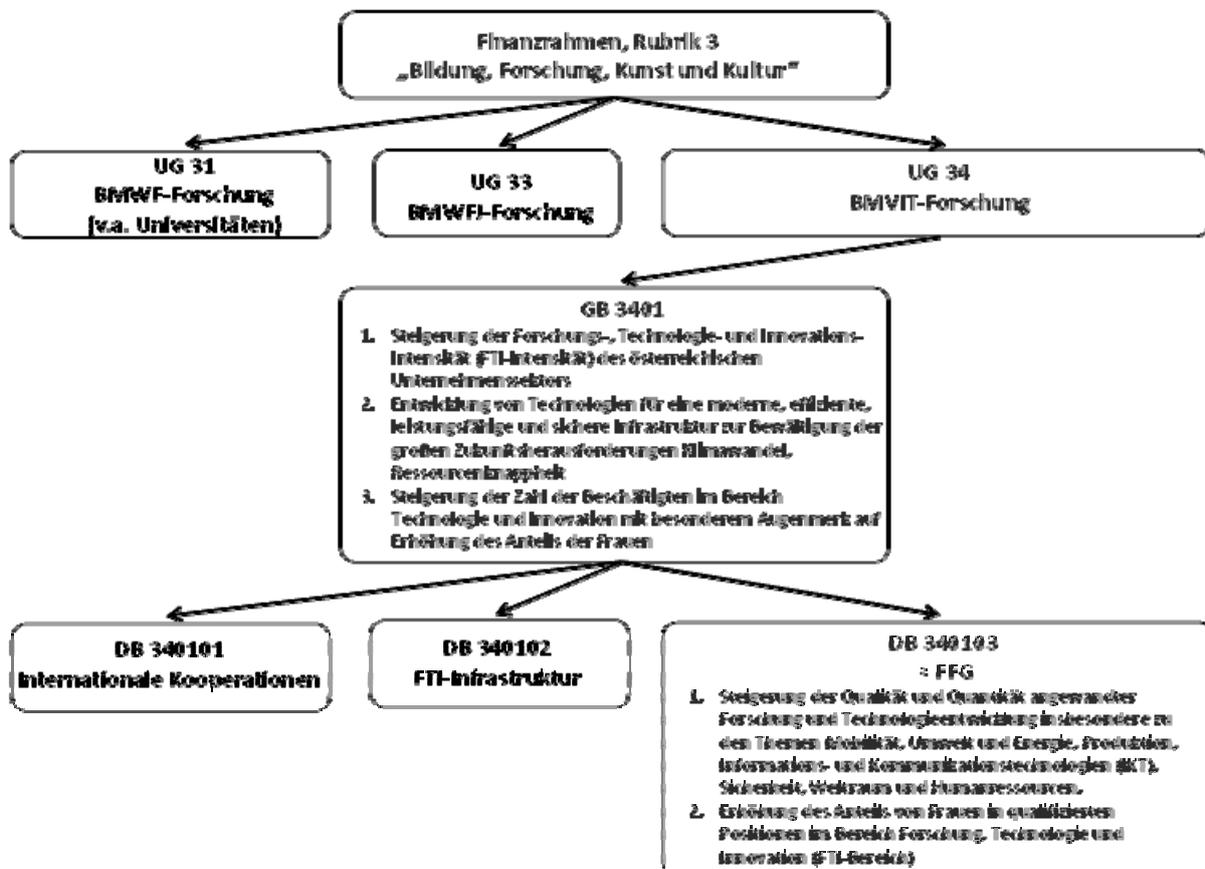
Im Rahmen der WHF definiert die Verwaltung in einem ersten Schritt Wirkungsziele, Maßnahmen zur Erreichung dieser Wirkungsziele sowie Indikatoren bzw. Kennzahlen, welche die Ziele operationalisieren. Weiterhin setzt sich die jeweilige Verwaltungseinheit auch numerisch bestimmte Ziele (Meilensteine), welche in einem jährlich stattfindenden Ist-Soll-Vergleich auf ihren Zielerreichungsgrad hin überprüft werden sollen. Das Monitoring der Zielerreichung (Wirkungscontrolling) erfolgt durch das Bundeskanzleramt. Materiell wirksame Sanktionen bei fehlender Zielerreichung sind bislang nicht vorgesehen. Zu beachten ist, dass Ziele entsprechend der in Abbildung 9 dargestellten Budgetgliederung des Finanzrahmens auf mehreren Ebenen definiert werden und einander in sachlogischer Hinsicht hierarchisch zugeordnet werden, also Ziele auf Ebene der Untergliederungen (UG) werden konkretisiert auf Ebene der Global- und Detailbudgets. Demnach hat die WHF wenig mit herkömmlichen, externen Evaluierungen im FTI-Bereich zu tun. Insbesondere ist festzustellen, dass es sich dabei um einen internen Verwaltungsprozess handelt, d.h. die mit der WHF verbundenen Aktivitäten erfolgen durch die Verwaltung selbst.

⁶ http://m.bmf.gv.at/Budget/Haushaltsrechtsreform/2EtappederHaushalts_10081/_start.htm [16.11.2011].

Um eine Einschätzung potentieller Folgen der WHF auf die FTI-Politik zu ermöglichen, ist zunächst ein kurzer Einblick in die Budgetstruktur nach dem neuen Bundesfinanzrahmengesetz notwendig (Abbildung 10). Demnach besteht das Budget aus 5 Rubriken, wobei die Rubrik 3 die für die FTI-Politik relevante ist. Vom finanziellen Volumen handelt es sich hierbei um die zweitgrößte Rubrik nach der mit Abstand größten Rubrik 2 (Arbeit, Soziales, Gesundheit und Familie). Diese Rubrik setzt sich aus 3 sogenannten Untergliederungen (UG) zusammen.

Die relevanten UGs für den F&E-Sektor sind UG 31, UG 33 und UG 34. Das UG 34 entspricht dabei dem Forschungsetat des BMVIT. Dem zugeordnet sind sogenannte Globalbudgets, die wiederum die Summe von Detailbudgets sind. Abbildung 10 macht deutlich, dass die WHF auf einer bereits relativ hoch aggregierten Ebene ansetzt. Während die Grobziele auf Ebene der Untergliederungen formuliert werden, bilden die Ziele auf Ebene der Detailbudgets die Feinziele ab. Betrachtet man die Detailbudgets, so stellt die FFG eines von drei Detailbudgets des Globalbudgets GB 3401 dar. Damit ist deutlich, dass die Wirkungsziele letztlich für die FFG insgesamt gelten und weder einzelne Programme noch einzelne Projekte unmittelbar zu steuern versuchen. Die Ziele auf Ebene des Globalbudgets sollen durch Aktivitäten in drei Detailbudgets erreicht werden, wobei neben der FFG auch internationale Kooperationen (DB 340101) (e.g. ESA) und FTI-Infrastruktur (DB 340102) zum GB 3401 zählen.

Abbildung 10: Budgetstruktur und WHF mit zugeordneten Wirkungszielen für den FTI-Sektor nach Entwürfen des BMVIT



Eigene Darstellung

Die in Abbildung 10 dargestellten Wirkungsziele auf Ebene der Global- und Detailbudgets sind als Entwürfe des BMVIT zu verstehen und stehen noch nicht endgültig fest. Auf Ebene der Untergliederungen enthält bereits der Strategiebericht zum Bundesfinanzrahmengesetz 2012-2015 folgende Hinweise auf Zielsetzungen (BMF 2011:14): „ Es erfolgt eine Bündelung der Mittel des BMVIT in folgenden vier Schwerpunkten: Intelligente Mobilität, Energie und Nachhaltigkeit, Informations- und Kommunikationstechnologie, Intelligente Produktion sowie Sicherheit kritischer Infrastruktur und Weltraum. (...) Die Allokation der öffentlichen Mittel soll einer verstärkten Output- und Wirkungsorientierung folgen.

Weiterhin soll verstärkt eine langfristige budgetäre Planungssicherheit durch mehrjährige Vereinbarungen auf Ressortebene hergestellt werden. Die Aufgabenverteilung zwischen den Ressorts und Förderagenturen soll durch höhere operative Unabhängigkeit der Agenturen und gleichzeitig verstärkte operative Steuerung durch die Ressorts optimiert werden.“ Nachdem der UG 34 nur ein Globalbudget (GB 3401) zugeordnet ist, entsprechen die Ziele der UG 34 demjenigen des GB 3401.

Eine Zusammenschau von geplanten Wirkungszielen, Maßnahmen und Indikatoren für das Global- und Detailbudget, welches für die FFG von Relevanz ist, zeigt Tabelle 8. Es handelt sich um drei Wirkungsziele auf Globalbudgetebene und zwei Wirkungsziele auf Detailbudgetebene. Deutlich wird hier auch die Bedeutung von Gender-Aspekten, ist doch ein Wirkungsziel auf Ebene des Detailbudgets explizit ein Ziel zur Förderung von Frauen im FTI-Sektor, während auch eines von drei Globalbudgetzielen eine Erhöhung des Beschäftigtenanteils der Frauen in F&E beinhaltet. Neben den Gender-Aspekten ist auch die zunehmende thematische Orientierung im Rahmen des Themenmanagements bereits deutlich erkennbar. Neben den beiden Schlüsselproblemen Klimawandel und Ressourcenknappheit werden die Themen Mobilität, Umwelt und Energie, Produktion, Informations- und Kommunikationstechnologien (IKT), Sicherheit, Weltraum und Humanressourcen genannt. Allerdings zeigen die vorgesehenen Indikatoren keine Auswertung nach diesen „Themen“ an.

Tabelle 8: Entwurf zu Zielen, Maßnahmen und Indikatoren zur WHF im BMVIT

	Ziele	Maßnahmen	Indikatoren
Globalbudget 3401	Steigerung der Forschungs-, Technologie- und Innovations- Intensität (FTI-Intensität) des österreichischen Unternehmenssektors	Förderung der unternehmensorientierten und außeruniversitären Forschung und Technologieentwicklung	Anzahl der Unternehmen im Sektor Forschung und experimentelle Entwicklung (F&E)
	Entwicklung von Technologien für eine moderne, effiziente, leistungsfähige und sichere Infrastruktur zur Bewältigung der großen Zukunftsherausforderungen Klimawandel, Ressourcenknappheit		Beschäftigte in Forschung und experimentelle Entwicklung (F&E) im Unternehmenssektor
	Steigerung der Zahl der Beschäftigten im Bereich Technologie und Innovation mit besonderem Augenmerk auf Erhöhung des Anteils der Frauen	Bevorzugte Förderung von Vorhaben der unternehmensorientierten und außeruniversitären Forschung und Technologieentwicklung nach genderspezifischen Kriterien	Frauenanteil von wissenschaftlichem Personal und höherqualifizierten-nicht-wissenschaftlichen Personal im Unternehmenssektor inklusive kooperativen Bereich im Unternehmenssektor
Detailbudget 340103		Förderung von unternehmensbezogenen Forschungsvorhaben	Anteil der Unternehmen, die ihre Forschungstätigkeit ausweiten
	Steigerung der Qualität und Quantität angewandter Forschung und Technologieentwicklung insbesondere zu den Themen Mobilität, Umwelt und Energie, Produktion, Informations- und Kommunikationstechnologien (IKT), Sicherheit, Weltraum und Humanressourcen.	Förderung der Kooperationen von Unternehmen, Forschungseinrichtungen und Bedarfsträgern	Anzahl von Unternehmen, die eine Forschungstätigkeit neu aufnehmen
		Qualitative Steigerung des FTI-Niveaus in Unternehmen	Anteil von Kooperationen Wiss-Wirtschaft an Gesamtprojekten
			Verbessertes technologisches Niveau im Unternehmen
			Verbessertes Projektmanagement und Innovationsmanagement
	Erhöhung des Anteils von Frauen in qualifizierten Positionen im Bereich Forschung, Technologie und Innovation (FTI-Bereich)	Differenzierte Beurteilung von Projekten (=explizite Angabe in Projekten) nach höherem Frauenanteil Erhöhung des Einflusses von Forscherinnen und Expertinnen im FTI-Bereich	Anteil von Projektleiterinnen in geförderten Projekten Anteil von Frauen in Bewertungsgremien (Jury)

Eigene Darstellung, Angaben BMVIT

Analysiert man die Ziele bzw. Indikatoren nach der Beeinflussbarkeit durch die Politik, so zeigt sich, dass letztlich nur drei von zehn Zielindikatoren (Anteil von Kooperationen Wissenschaft-Wirtschaft an Gesamtprojekten, Anteil von Projektleiterinnen in geförderten Projekten, Anteil von Projektleiterinnen in geförderten Projekten) unmittelbar einer Steuerung durch die FFG unterliegt. Alle anderen Indikatoren werden durch eine Vielzahl von Faktoren beeinflusst, wobei der Faktor „FTI-Politik“ wohl nur einen relativ bescheidenen Einfluss auf derart hoch aggregierte Größen ausüben dürfte. Damit stellt sich die Frage, wie ein Zusammenhang zwischen der Entwicklung der Indikatoren und der FTI-Politik hergestellt werden soll.

Die Identifikation der Effekte FTI-politischer Intervention basiert auf einem Vorher-Nachher-Vergleich. Neben den Problemen einer solchen Methodik, würde dies der Tendenz nach dazu führen, alle Veränderungen des Innovationssystems im Sinne der in Tabelle 8 dargestellten Indikatoren, auf politische Intervention zurückzuführen. Nachdem dies insbesondere im Aggregat (e.g. Anzahl der Unternehmen im Sektor Forschung und experimentelle Entwicklung (F&E)) durchwegs falsch sein dürfte, stellt sich das Problem, welcher Anteil der Entwicklung FTI-politisch verursacht worden ist. Letztlich muss festgestellt werden, dass diese Frage nicht mit deskriptiven Methoden beantwortet werden kann. Wie noch zu zeigen sein wird, kann ein negativer Verlauf des Aggregats mit positiven Maßnahmeneffekten verbunden sein und vice versa.

Unmittelbar dürfte es gar keine Auswirkungen der WHF geben, es könnte aber eine Reihe mittelbarer Auswirkungen haben:

- Zunächst einmal kommt es zu einer Aufwertung des Wirkungsmonitoring, welches aktuell auf alle Programme der FFG ausgeweitet wird. Aus methodischer Perspektive ist dies zumindest fragwürdig, weil hier Unternehmen direkt über Programmwirkungen gefragt werden (siehe Kapitel 5.2).
- Mittelfristig könnte es auch zu einer partiellen Re-orientierung von Interim- und Ex-post Evaluierungen kommen. Dieses Argument ist vor dem Hintergrund einer Zurechenbarkeit politischer Maßnahmen auf die genannten Indikatoren zu sehen. Nachdem dies, wie erwähnt, mit deskriptiven Methoden letztlich nicht bewerkstelligt werden kann, könnte die Nachfrage nach methodisch effektiveren Evaluierungen zunehmen. Diese Entwicklung lässt sich jedenfalls in den USA im Gefolge der Einführung des „Government Research and Performance Act“ von 1993 beobachten (Link, Scott 2011). Weiterer Faktoren in diesem Zusammenhang betreffen die steigende Betonung der Output- und Wirkungsorientierung zur Allokation öffentlicher F&E-Subventionen durch die FTI-Strategie des Bundes sowie die Betrachtung von FTI-Interventionen aus einer Effizienzperspektive.
- Einen größeren Einfluss wird die Berücksichtigung von Fragen der Gendergerechtigkeit in Evaluierungen spielen. Diese Annahme wird bereits in den Wirkungszielen reflektiert, wo der Gender-Aspekt eine prominente Rolle einnimmt.

- Eine weitere Auswirkung könnte auf Ebene der ex-ante Evaluierungen, d.h. bei der Projektselektion, eintreten. Eine Steuerung von Programmen oder Institutionen mittels Indikatoren verändert in der Regel die Anreize zumindest derjenigen Akteure, welche diese Ziele erfüllen sollen (Gibbons 1998). Während dies im Fall intendierter Verhaltensänderungen freilich erwünscht ist, zeigt die Empirie eine Reihe von Beispielen, für nicht-intendierte, adverse Verhaltensreaktionen durch kennzahlenbasierte Steuerung.
- Tabelle 8 zeigt jedoch, dass die vorgesehenen Indikatoren der WHZ aufgrund ihrer Allgemeinheit und Breite keine negativen Effekte haben sollten. Trotzdem wurde in den Interviews die Frage aufgeworfen, ob es nicht aufgrund einer forcierten Wirkungsorientierung zur weiteren Abnahme der Risikobereitschaft im Rahmen der Projektselektion kommen könnte. Auch die relativ kurzen Zeiträume zur Zielerreichung (1 Jahr) könnten zu dieser Entwicklung beitragen. Bedenkt man die häufig wiederholte Forderung nach einer stärkeren Risikoneigung der direkten unternehmensbezogenen Forschungsförderung (Aiginger 2009, Schibany 2009), so hätte dies tatsächlich negative Konsequenzen für das österreichische Innovationssystem. Demgegenüber steht das Argument, dass ein Einfluss auf die Projektselektion höchst unwahrscheinlich ist, da die Wirkungsziele eben nicht auf Programmebene sondern eher auf Institutionenebene definiert sind. Dieses Argument basiert jedoch auf der kritischen Annahme, dass übergeordnete Zielsetzungen keinen Einfluss auf das Mikromanagement haben. Aus Sicht der StudienautorInnen bedeutsamer ist die inhaltliche Ausgestaltung der Ziele. Diese scheinen jedenfalls auf den ersten Blick risikoneutraler Natur zu sein. Eher problematisch ist möglicherweise die jährliche Zielerreichung. Werden vermehrt radikale Innovationen gefördert, so ist auch mit einer erhöhten Volatilität von Outputvariablen zu rechnen. Dies würde die Wahrscheinlichkeit zur Verfehlung der Ziele zwar nicht im Erwartungswert, aber auf jährlicher Basis erhöhen.

Die Diskussion um den "Government Research and Performance Act" brachte jedenfalls eine Änderung hinsichtlich des Zeitraums, innerhalb dessen Ziele erreicht werden sollen: „The initial focus on short-term activity indicators has given way to a call for longer-term processes that develop a more strategic view.“ (Cozzens 1999). Es stellt sich aber auch die Frage, ob es nicht sinnvoll wäre, Ziele zu definieren, die explizit die Förderung von Risiko zum erstrebenswerten Ziel erklären. Generell kann jedoch über mögliche Verhaltensänderungen ohne klare Informationen über Konsequenzen bei Zielverfehlung nur gemutmaßt werden.

- Last but not least wird das Finanzministerium weiter an Einfluss auf die FTI-Politik gewinnen. Auch dieser Trend ist keineswegs ein österreichisches Spezifikum, sondern ein Trend, der in den OECD-Staaten seit einigen Jahrzehnten zu beobachten ist (Klump 2006). Durch die Finanz- und Wirtschaftskrise wird dieser Trend noch weiter beschleunigt.

4 Nationale und internationale Evaluierungstrends von innovationspolitischen Maßnahmen im Überblick

„Das Vergleichen ist das Ende des Glücks und der Anfang der Unzufriedenheit.“

Sören Kierkegaard (1813-55)

Kapitel 4 bietet einen Überblick über nationale und internationale Evaluierungstrends. Aktuelle Entwicklungen innovationspolitischer Evaluation in Österreich zeigt Abschnitt 4.1. Daran anschließend folgt ein internationaler Vergleich österreichischer Entwicklungen mit einigen ausgewählten Beispielen aus anderen OECD-Staaten. In Abschnitt 4.2. werden auf der Basis einer detaillierten Analyse von 10 aktuellen Programmdokumenten und Evaluierungen eine Reihe von Empfehlungen zu unterschiedlichen Aspekten von FTI-Programmen und Evaluierungen abgeleitet. Im internationalen Vergleich (Abschnitt 4.3.) zeigt sich eine signifikant geringere Methodenvielfalt und eine Lücke beim Einsatz analytischer, quantitativer Methoden in Österreich relativ zu anderen OECD-Staaten.

4.1. AUSGEWÄHLTE ÖSTERREICHISCHEN EVALUIERUNGEN VON 2005-2011

Die Wahrnehmung von Wirkungsanalysen FTI-politischer Programme hat sich in den letzten Jahren stark gewandelt. Vormalig noch als „starre“ und von „außen“ angeordnete Überprüfungen verstanden, steht nun bei Wirkungsanalysen das „Verstehen“ von Politikprozessen im Vordergrund. Damit geht auch ein Wandel der methodischen Tools einher: der Fokus liegt nun auf „flexiblen“, „zielgerichteten“ und „kombinierten“ Verfahren.

Zu den österreichischen Standardwerkzeugen gehören die standardisierte direkte Befragung, qualitative Interviews sowie Fallstudien. Die meisten österreichischen Wirkungsanalysen verwenden einen „Methodenmix“ (Kombination von qualitativen und quantitativen Verfahren) um das Verständnis über Politikprozesse und Wirkungszusammenhänge zu erhöhen. Dabei dominieren eindeutig qualitative Verfahren. Soziale Netzwerkanalysen sowie aktuelle ökonometrische Entwicklungen wie Matched Pairs, Differences-in-Differences Estimators oder Sample Selection Estimators werden, zumeist aufgrund von Datenbeschränkungen, nicht angewendet.

Nachfolgende Tabelle gibt einen Überblick über häufig verwendete Methoden in FTI-Evaluationsstudien aus den Jahren 2005-2011.

Tabelle 9: Österreichische FTI-Evaluierungen 2005-2011 im Überblick

Titel	Jahr	Evaluations- typ	Auftrag- geber	Auftrag- nehmer	Inhaltlicher Fokus des/der evaluierten Programms/Institution	Methoden
FIT-IT Evaluierung	2005	Interims- evaluierung	BMVIT	Joanneum Research	Entwicklung signifikanter technologischer Innovationen	– Desk Research/Dokumentenanalyse – Schriftliche, standardisierte Befrag.
GENAU Evaluierung	2005	Programm- Management- evaluierung	BMBWK	Joanneum Research, TIA Consulting, KMU Forschung Austria	Stärkung und Vernetzung der Genomforschung in Österreich	- Desk Research/Dokumentenanalyse - Logic Chart Analyse - Qualitative Interviews - Case Studies - Social Network Analysis
FWF Evaluierung	2005	Ex-Post und Performance Evaluierung	FWF	Joanneum Research	Einrichtung zur Förderung exzellenter Grundlagenforschung	- Literaturrecherche - Desk Research/Dokumentenanalyse - Deskriptive statistische Auswertungen (FWF-Datenbank) - Multiple lineare Regressionsanalysen
NANO Evaluierung	2006	Interims- evaluierung	BMVIT	Technopolis, NMTC	Förderung der Forschung und Technologieentwicklung in den Nanowissenschaften und Nanotechnologien	- Workshops/Hintergrundgespräche - Vorstrukturierte Interviews - Offene Interviews (persönlich+telefonisch) - Standardisierte, schriftliche online- Befragung
Basis- Programme Evaluierung	2006	Projekt- Evaluierung (Effizienz- analyse)	FFG	KMU Forschung Austria	Förderung von technisch schwierigen und riskanten FuEuI Projekten mit erkennbarem Potenzial zur wirtschaftlichen Verwertung	- Desk Research/Dokumentenanalyse - Standardisierte, schriftliche Befragung
AplusB Evaluierung	2008	Interims- evaluierung	BMVIT	Inno Germany AG	Nachhaltige Steigerung sowohl der Anzahl als auch des Erfolgs von Gründungen aus Hochschulen und außeruniversitären Forschungseinrichtungen	- Desk Research/Dokumentenanalyse - Deskriptive Auswertungen, FFG Datenbank - Qualitative Leitfadenterviews (persönlich + telefonisch) - Internationaler Workshop

Titel	Jahr	Evaluations- typ	Auftrag- geber	Auftrag- nehmer	Inhaltlicher Fokus des/der evaluierten Programms/Institution	Methoden
Bridge Evaluierung	2009	Programm- evaluierung	BMVIT	Technopolis	Weiterentwicklung und Nutzbarmachung von Erkenntnissen der Grundlagenforschung für wirtschaftliche Anwendungen	<ul style="list-style-type: none"> - Desk Research/Dokumentenanalyse - Deskriptive Auswertung der Monitoring Daten, FFG - Schriftliche, standardisierte online- Befragung - Qualitative Interviews - Case Studies
System Evaluierung	2009	System- Evaluierung	BMVIT, BMWFJ	WIFO, prognos, OeNB, convelop, KMU Forschung Austria	Österreichisches Forschungs-, Technologie- und Innovationssystem	<ul style="list-style-type: none"> - Literaturrecherche - Dokumentenanalysen - Quantitative, statistische Auswertungen (Sekundärdaten – EUROSTAT, Statistik Austria)
TAKE OFF Evaluierung	2009	Interims- evaluierung	BMVIT	prognos	Koordination und Ausbau des Luftfahrtsektors als Gesamtsystem	<ul style="list-style-type: none"> - Auswertung der Monitoringdaten, FFG - Standardisierte, schriftliche Befragung - Qualitative Interviews - Peer Reviews - Soziale Netzwerkanalyse
AplusB Evaluierung	2010	Ex-post Evaluierung (?)	FFG	Quadris Consulting GmbH	Nachhaltige Steigerung sowohl der Anzahl als auch des Erfolgs von Gründungen aus Hochschulen und außeruniversitären Forschungseinrichtungen	<ul style="list-style-type: none"> - Dokumentenanalyse - Statistische Auswertung der Monitoringdaten, FFG - Kontrollgruppenansatz
Laura-Bassi Centres of Excellence	2011	Interims- evaluierung	BMWFJ	KMU Forschung Austria	Förderung wissenschaftliche Exzellenz, Gleichstellung und Managementaspekte im naturwissenschaftlichen Forschungsbereichen (?)	<ul style="list-style-type: none"> - Literaturrecherche - Dokumentenanalyse - Standardisierte, schriftliche Befragung - Qualitative Interviews (persönlich + telefonisch) - Fokusgruppen - Workshops

Titel	Jahr	Evaluations- typ	Auftrag- geber	Auftrag- nehmer	Inhaltlicher Fokus des/der evaluierten Programms/Institution	Methoden
Headquarter Programm- evaluierung	2011	Programm- evaluierung	BMVIT	Technopolis	Nachhaltigen Auf- und Ausbau neuer Forschungs- und Entwicklungskompetenz	<ul style="list-style-type: none"> - Literaturrecherche - Dokumentenanalyse - Auswertung der Förderdaten - Qualitative Leitfadeninterviews (persönlich + telefonisch)
Benefit Evaluierung	2011	Portfolio- und Wirkungs- analyse	BMVIT	KMU Forschung Austria	Förderung von Forschung und Entwicklung für innovative, Technologie-Produkte (insbesondere im IKT-Bereich) und technologiegestützte Dienstleistungen	<ul style="list-style-type: none"> - Literaturrecherche - Dokumentenanalyse - Logic Chart Analyse - Analyse sekundärstatistischer Daten, FFG - Qualitative Leitfadeninterviews (persönlich + telefonisch) - ExpertInneninterviews - Standardisierte, schriftliche online-Befragung

4.2. ERGEBNISSE AUS DER ANALYSE DER PROGRAMMDOKUMENTE UND EVALUATIONEN

In diesem Kapitel werden die Ergebnisse und Empfehlungen präsentiert, die auf Basis einer Durchsicht und Analyse der Programmdokumente und Evaluationen von zehn verschiedenen Programmen des BMVIT/FFG formuliert wurden. Nachfolgende Tabelle listet die analysierten Programme auf:

Tabelle 10. Übersicht der analysierten Programmevaluationen

Programm	Evaluation	AutorInnen
FIT-IT	Interimsevaluierung, April 2005	Klaus Zinöcker, Julia Schindler, Wolfgang Polt, Stefanie Gude, Michael Stampfer, Julia Schmidmayer
NANO	Interimsevaluierung, November 2006	Leonhard Jörg, Matthias Werner
GENAU	Programmmanagement Evaluierung, Juni 2005	Klaus Zinöcker, Brigitte Tempelmeier, Alfred Radauer, Roald Steiner, Rosalie Ruegg, Iris Fischl, Franziska Steyrer, Helene Schiffbänker, Jaqueline Allan
COIN	Programmdokument	
KIRAS	Programmdokument	
APLUSB	Evaluierung, Sept. 2010	Kirsten Tangemann, Siegfried Vössner
BENEFIT	Portfolio- und Wirkungsanalyse, 2011	Aliette Dörflinger, Sascha Ruhland,
BRIDGE	Programmevaluierung, Juli 2009	Katharina Warta, Barbara Good, Anton Geyer
TAKE OFF	Zwischenevaluierung, 2009	Oliver Pfirrmann, Stephan Heinrich, Thomas Herzog
FEMTECH	Programmevaluierung, Main 2011	Karin Grasenick, Gabriele Gerhardtter, Nicole Warthun, Andrea Löther, Markus Gabriel, Nina Steinweg, Stephan Kupsa, Erick Domke, Martina Süßmayer
HEADQUARTER	Programmevaluierung, März 2011	Anton Geyer, Brigitte Tiefenthaler

Die Programmdokumente und Evaluationen wurden in Bezug auf folgende Aspekte analysiert:

1. Ziele des Programms
2. Daten, die in den Evaluationen verwendet werden
3. Evaluationsmethoden
4. Sozioökonomische Wirkungsdimensionen
5. Indikatoren.

4.2.1 Ziele

Die exakte Formulierung von klaren und eindeutigen Zielen in den Programmdokumenten der verschiedenen FTI-Programme ist für aussagekräftige Ergebnisse und Schlussfolgerungen über die Zielerreichung und die festgestellten Wirkungen eines Programms von Bedeutung.

Die Analyse der Programmdokumente zeigt, dass dies in der Regel sehr gut gelingt. Die analysierten Dokumente weisen schlüssige, umfassende und klar mit der Mission des Programms verknüpfte Ziele auf.

Programme, insbesondere thematische Programme, die zusätzlich zu FTI-politischen Zielsetzungen soziale, nachhaltigkeitsbezogene, umweltbezogene und energiepolitische Dimensionen aufnehmen (z.B. KIRAS, Benefit, NANO) stellen EvaluatorInnen vor große methodische Herausforderungen bei der Messung dieser Ziele.

Um eine mögliche „Zielüberfrachtung“ dieser Programme (insbesondere KIRAS) zu vermeiden, wird empfohlen, eine Priorisierung und Hierarchisierung der Ziele vorzunehmen.

Weiters wird empfohlen, eine Einteilung der Ziele auf verschiedenen Zielebenen einheitlich für alle Programme durchzuführen. Das Hauptunterscheidungsmerkmal für die verschiedenen Zielebenen sollte die „Steuerbarkeit“ durch das Programm selbst sein. Eine mögliche Einteilung der Ziele könnte folgendermaßen aussehen:

Box 5: Beispiel für eine Einteilung FTI-politischer Zielsetzungen auf verschiedene Zielebenen

1. „Leitziele“ oder „Missionsziele“ des Programms

z.B. Erhöhung der Wettbewerbsfähigkeit Österreichs, Erhöhung des Frauenanteils im FTI-Bereich, Erhöhung der Lebensqualität älterer Menschen; diese kennzeichnen sich dadurch, dass sie langfristig, „global“, und zumeist nicht direkt messbar sind.

2. „Mittlerziele“ oder spezifische Ziele“

z.B. Erhöhung der Attraktivität technischer Berufe für Frauen, Bewusstsein für technische Neuerungen bei älteren Menschen etc.; diese werden aus den Leitzielen abgeleitet, sind mittelfristig und geben Handlungsansätze vor, sie sind jedoch oftmals nicht hinreichend genau messbar.

3. „Handlungsziele“ oder „operationale Ziele“

z.B. Frauen in FTI und ihr Potential sichtbar machen, Informationen zum Thema „technische Erneuerungen“ für die Zielgruppe ältere Menschen aufbereiten und zugänglich machen etc.; diese werden aus den Leit- und Mittlerzielen abgeleitet, sind klar operationalisierbar, direkt durch die FördernehmerInnen steuerbar und sollten zu einem vorbestimmten Zeitpunkt erreicht werden können.

Bei der Definition der Ziele und Strategien eines Programms sollte weiters spezifisch darauf geachtet werden, dass die Ziele eines Programms klar von den Strategien (= Handlungsanleitungen um diese Ziele zu erreichen) getrennt werden, um eine exakte Messung der Ziele durch das Evaluationsteam möglich zu machen.

Bei FemTech kam es teilweise zu einer Vermischung von Zielen und Strategien; hierzu der Kommentar des Evaluationsteams:

„In den Dokumenten von FemTech werden die Ziele in allgemeine, explizite und spezielle Ziele unterteilt. Dabei werden jedoch Ziele und Strategien zum Teil nicht klar abgegrenzt. So wird beispielsweise das Ziel der Bewusstseinsbildung, Sensibilisierung und Sichtbarmachung von Frauen in FTI als Strategie dargestellt, um Veränderungen in den Strukturen und Kulturen von Forschung und Technologie zu erreichen. Gleichzeitig lassen sich die Bereiche „Bewusstseinsbildung“ und „Veränderung von Strukturen“ jedoch auch als zwei gleichrangige Ziele identifizieren, die durch den Bereich „Förderung“ (Programmlinie Karriere) und „Aktivitäten“ innerhalb von FemTech verfolgt werden sollen (S. 19).“

Mit der Ausweitung des FTI-politischen Zielkatalogs um breitere gesellschaftspolitische und soziale Zielsetzungen, wäre es auch interessant, bereits in die Programmdokumente Ziele zu integrieren, die prozesshafte und verhaltensbezogene Änderungen durch das Programm selbst abfragen. Qualitativ angelegte Verfahren sind aktuell ein vielversprechender Weg, um Hinweise auf mögliche Wirkungen zu erhalten.

4.2.2 Datenlage

Der Zugang zu den entsprechenden Daten sowie ihre Qualität bestimmen zu einem großen Teil den Einsatz bestimmter Evaluationsmethoden und damit die Robustheit der Ergebnisse, die Schlussfolgerungen sowie die Aussagekraft der Evaluationen.

Aufgrund von Datenmangel haben quantitative kausale Wirkungsabschätzungen und experimentelle Methoden in Österreich Seltenheitswert; um auf Veränderungen durch das Programm schließen zu können, stützt man sich – so zeigt der Großteil der analysierten Evaluationen – auf von den FördernehmerInnen selbst wahrgenommene Veränderungen. Dies bringt erhebliche „Response-Probleme“ (soziale Erwünschtheit, Nicht-Wissen, Akquieszenz) mit sich und schränkt die Güte und Robustheit der Ergebnisse deutlich ein.

Hinsichtlich der Datenlage in Österreich ergibt sich daher ein deutlicher Verbesserungsbedarf in den folgenden Bereichen:

Zugang zu neuen Daten

Um quantitative kausale Wirkungsabschätzungen zu ermöglichen, ist eine sogenannte „Nullmessung“ der FördernehmerInnen vor Programmstart unabdingbar. Diese Nullmessung ermöglicht eine exakte Einschätzung der ökonomischen und sozialen Lage z.B. eines Unternehmens vor Programmstart. Im Vergleich mit einer „Kontrollgruppe“ an Unternehmen und den jeweiligen Messungen nach Ende des Programms können Programmwirkungen zugewiesen und quantifiziert werden. Bereits beim Programmdesign muss daher die Datengenerierung für den Zweck der Evaluation mitbedacht werden.

Zugang zu existierenden Daten

Die Statistik Austria besitzt Daten (CIS-Daten, F&E Erhebung), die für die Anwendung quantitativer analytischer Evaluationsmethoden bedeutsam sind; EvaluatorInnen können jedoch aktuell nicht – in der Regel aus Gründen des Datenschutzes sowie der erforderlichen Mittel – mit diesen Daten arbeiten. Dieser Punkt wird in Kapitel 7 ausführlich und im internationalen Vergleich erörtert. Hier besteht dringend der Bedarf, bestehende Datensätze der FFG und der Statistik Austria miteinander zu verknüpfen, um die Qualität und Güte kommender Evaluationen zu sichern.

Qualität existierender Daten

Bestehende Daten/Monitoringdaten müssen regelmäßig auf ihre Qualität und Aktualität geprüft werden. In der Headquarter-Evaluierung findet sich ein Kommentar des Evaluationsteams, in welchem die Seriosität der Unternehmensangaben angezweifelt wird: „In den Headquarter-Zusatzberichten müssen die Unternehmen Angaben über die FuE-Personaleffekte der Förderung machen, insbesondere wie viele neue FuE-Stellen im Rahmen der Headquarter Projekte geschaffen wurden. Eine seriöse Interpretation der Unternehmensangaben sowie der Nachhaltigkeit des FuE Mitarbeiteraufbaus ist jedoch unserer Einschätzung nach nicht möglich, da insbesondere über die Entwicklung nach Ende der Projekte kaum Informationen vorliegen. Wir verzichten daher auf eine

Zusammenfassung der Angaben der Unternehmen in den HQ-Zusatzberichten“ (Geyer, Tiefenthaler 2011).

Die Qualität der Unternehmensangaben muss verbessert werden, um Evaluationsteams auch quantitative Indikatoren für ihre Analyse zur Verfügung zu stellen.

4.2.3 Evaluierungsmethoden

Die Wahl der Evaluationsmethoden ist von der Datenverfügbarkeit und dem Evaluationsbudget abhängig und muss an den jeweiligen Evaluationszweck und Evaluationsgegenstand angepasst werden. Die große Bandbreite FTI-politischer Zielsetzungen erfordert oft eine Kombination verschiedener quantitativer und qualitativer Verfahren, die, miteinander verknüpft, das Wissen über die soziale und ökonomische Realität erhöhen.

Die Analyse der Dokumente ergab, dass die standardisierte direkte Befragung der FördernehmerInnen sowie qualitative Interviews, ExpertInnengruppen und Case Studies die „Standardwerkzeuge“ in den Evaluationen darstellen.

Vor diesem Hintergrund können folgende Empfehlungen genannt werden:

Bedarf an quantitativen analytischen Methoden

Quantitative Evaluationsmethoden, soziale Experimente, Kosten-Nutzen Analysen sowie aktuelle Methoden der ökonometrischen Forschung, wie etwa matched pairs, differences in differences estimators oder natural experiments, kommen in der österreichischen evaluatorischen Praxis sehr selten vor, bieten aber den Vorteil quantifizierbarer Wirkungsabschätzungen. Auch ex-ante Simulationsmodelle wie z.B. die agentenbasierte Modellierung wurde noch nicht für österreichische FTI-politische Wirkungsanalysen eingesetzt. Der Grund dafür ist die mangelnde Datenlage.

Verstärkter Einbezug innovativer, partizipativer Verfahren

Die Öffnung des FTI-Bereichs gegenüber sozialen und gesellschaftspolitischen Zielsetzungen erfordert ein neues Methodenspektrum. Case Studies, ExpertInnen-Interviews und qualitative Interviews sind ein vielversprechender Weg bei der Analyse dieser Themenbereiche und kommen in den genannten Evaluationen bereits häufig zum Einsatz. International besteht zusätzlich der Trend, insbesondere bei der Evaluierung nicht-ökonomischer Wirkungsdimensionen partizipative Methoden zu verwenden (e.g. participatory social mapping, causal-linkage and change and trend diagramming etc). Die Stärke von partizipativen Methoden ist es, die „Öffentlichkeit“, relevante „stakeholder“ oder „enduser“ in den Evaluationsprozess direkt miteinzubeziehen. Dabei geht es explizit nicht um geförderte Unternehmen. Diese Methoden würden sich insbesondere für thematische Programme wie Benefit, Kiras, FemTech oder GENAU anbieten, wurden bislang aber noch nicht angewendet.

Monetäre Bewertung nicht-ökonomischer Effekte

Internationale Evaluationsstudien zeigen zahlreiche methodische Entwicklungen im Bereich der Messung des Social Return on Investments (SROI). Es geht dabei darum, für nicht-ökonomische Effekte (insbesondere Umwelteffekte, Gesundheitseffekte), für die kein adäquater Marktpreis existiert, eine ökonomische Bewertung zu finden.

Mithilfe von „stated- or revealed preference“- Methoden kann dies gelingen.

Hier lautet die Empfehlung, insbesondere für die thematischen Programme des BMVIT, ein erhöhtes Evaluierungsbudget in Kauf zu nehmen, um diese Methoden zu testen und weiterentwickeln zu können.

Einsatz von kreativen, modifizierten Fokusgruppen & quanti-quali Designs

Die Durchsicht der Evaluationen zeigt, dass Fokusgruppen in FTI-Evaluationen zum Einsatz kommen. Es besteht jedoch der Trend, vorwiegend das Instrument der „traditionellen Fokusgruppe“ zu verwenden; kreative, modifizierte Fokusgruppen, die Probleme der traditionellen Fokusgruppenforschung (group conformity, preference toward agreement etc.) zu reduzieren versuchen, haben noch kaum Eingang in das Feld österreichischer FTI-Evaluationen gefunden. Auch qualitative Fokusgruppendifkussionen, die quantitative Elemente enthalten, kommen kaum zum Einsatz. Diese Instrumente sollten verstärkt angewandt und weiterentwickelt werden. Bei der Auswertung von Fokusgruppenergebnissen ist auf die Analyse der Kontextfaktoren zu achten.

Sozioökonomische Wirkungsdimensionen

International stehen die oft sehr umfassenden FTI-relevanten Programmziele einem deutlich engeren Spektrum an Evaluierungskriterien von FTI-Programmen gegenüber, die ökonomische, technologische und wissenschaftliche Wirkungen messen sollen.

Auch die Analyse der österreichischen Dokumente zeigt, dass in den ausgewählten Programmen ökonomische, technologische und wissenschaftliche Wirkungen dominieren (Abbildung 11).

Evaluierungen erfolgreich durchführen zu können, ist die Bereitstellung höherer Evaluierungsbudgets erforderlich.

Indikatoren

Indikatoren dienen als „Anzeiger“ nicht-direkt beobachtbarer Konstrukte. Ein Indikator gestattet nur mit einer gewissen Wahrscheinlichkeit, auf das vorausgesetzte Indikaturniveau zu schließen. Insbesondere bei der Messung sozialer und gesellschaftspolitischer Zielsetzungen in FTI-Programmen ist die Wahl der passenden Indikatoren eine besondere Herausforderung.

Die Veränderungen/Wirkungen von nicht-ökonomischen Wirkungsdimensionen werden in den genannten Evaluationen häufig durch direkte schriftliche Befragungen ermittelt. Als Indikator für z.B. strukturelle Verbesserungen für Frauen im FTI-Bereich dient die Frage: „Wurden strukturelle Verbesserungen für Frauen durchgeführt? Stimme zu – lehne ab; siehe Programm FemTech). Die EvaluatorenInnen stützen sich bei ihren Einschätzungen damit auf die wahrgenommene Veränderung der FördernehmerInnen selbst. Dies bringt jedoch erhebliche methodische Probleme mit sich (soziale Erwünschtheit, Non-Response, Nicht-Wissen etc.)

Zu empfehlen wäre hier:

- **Einsatz verschiedener „objektiver“ Indikatoren:**

Die „subjektive“ Einschätzung hinsichtlich der wahrgenommenen Veränderungen könnte durch eine Reihe an „objektiv“ feststellbaren Aspekten ersetzt werden, beispielsweise:

- Sind Sie hierarchisch aufgestiegen?
- Haben Sie nun mehr Verantwortung?
- Verdienen Sie mehr?
- Um wie viele Personen hat sich ihr Netzwerk erweitert?

Durch diese Fragestellungen wird subjektiven Einschätzungen weniger Raum gegeben.

- **Einsatz unterschiedlicher Indikatoren zur Messung ein- und desselben Konstrukts**

Bei komplexen Sachverhalten (wie z.B. strukturellen Veränderungen, aber auch bei Konzepten wie Chancengleichheit oder Lebensqualität) sollte nicht nur eine Frage zum jeweiligen Thema gestellt werden, sondern mehrere Indikatoren eingesetzt werden, die auf unterschiedliche Aspekte ein- und desselben Konstrukts abzielen. Zeigen alle Indikatoren in dieselbe Richtung, so erhält man ein robusteres Ergebnis in Bezug auf die potentielle Wirkung einer Maßnahme, als wenn nur ein Indikator erhoben worden wird. Indikatoren sind zumeist nicht „universell“ anwendbar, sondern müssen in Bezug auf bestimmte Kontexte (national, lokal, individuell, aggregiert, etc.) adaptiert und kontinuierlich in Bezug auf ihre statische Validität und Reliabilität getestet werden. Vorsicht ist auch bei der Gewichtung und Anwendung zusammengesetzter Indikatoren geboten – es empfiehlt sich, die einzelnen Dimensionen, aus denen sich ein gewichteter Index zusammensetzt, einzeln aufzuschlüsseln, auch nach Geschlecht oder anderen interessierenden sozio-ökonomischen Charakteristika.

- **Verwendung von nominalen und ordinalen Indikatoren**

Es empfiehlt sich, Veränderungen nicht-ökonomischer Wirkungsdimensionen auf zwei verschiedene Arten zu erheben:

a.) als nominale Information (z.B. Erhöhte sich die Anzahl an Publikationen? Ja/nein)

Dieser Indikator macht es möglich, die Häufigkeit und die Art der Veränderung zu erfassen.

b.) als ordinale/numerische Information (z.B. wie viele Publikationen wurden veröffentlicht? Anzahl an Publikationen). Dieser Indikator zeigt die Bedeutung/Stärke der Veränderung an.

- **Indikatoren nicht nur für erwünschte Effekte bilden, sondern auch für unerwünschte Wirkungen**

In den analysierten Evaluationen ist auffallend, dass nur auf erwünschte Effekte des Programms getestet wird. Gerade aber bei nicht-ökonomischen Wirkungsdimensionen ist eine Testung auf unerwünschte Effekte aufgrund der hohen Komplexität (unterschiedliche Wirkungen – intendierte/nicht-intendierte – bei verschiedenen TeilnehmerInnen) der Dimensionen unabdingbar.

Beispielsweise könnte folgende Frage bei der FemTech-Evaluierung gestellt werden: Hat die Maßnahme „Expertin des Monats“ wirklich zu einer positiven und verstärkten Sichtbarkeit von Frauen in FTI geführt oder wurden die Forscherinnen, die diese Auszeichnungen bekamen, von der breiten wissenschaftlichen Community belächelt? Hat die Maßnahme Frauen gefördert oder wurde sie als positive Diskriminierung aufgefasst?

4.3. EVALUATION INNOVATIONSPOLITISCHER MAßNAHMEN IM INTERNATIONALEN VERGLEICH⁷

Ein internationaler Vergleich von Evaluationen im FTI-Bereich ist interessant und schwierig zugleich. Interessant ist der Vergleich, weil in der einschlägigen Literatur die Meinung vertreten wird, dass große Unterschiede zwischen den Staaten bestehen (Box 6).

⁷ Andere Abschnitte in diesem Bericht die sich explizit mit dem Thema internationaler Vergleich auseinandersetzen sind 2.1.2 sowie insbesondere Kapitel 5, welches Datenverfügbarkeiten in mehreren OECD-Staaten mit der Situation in Österreich vergleicht.

Box 6: Internationale Unterschiede bei der Evaluation von FTI-Politik

“Traditions and cultures of accountability have evolved differently in different national settings and influence performance assessment measures and processes. (...) Despite a trend towards convergence inspired by international experience, evaluation practices vary strongly across countries and significant obstacles to international learning remain.”

OECD (2010: 202, 304)

While it seems that different countries have established their own individual modes of conducting research and providing policy advice, the experiences made with these approaches are usually not compared systematically. This is all the more remarkable, as academic research in economics is conducted in an integrated world market.

Schmidt (2007:6)

Schwierig ist die Aufgabe eines internationalen Vergleichs aus zwei Gründen: Erstens sind nach wie vor viele Evaluierungen in der jeweiligen Nationalsprache verfasst. Zweitens sind Evaluierungen entweder gar nicht veröffentlicht oder einfach nicht an einer zentralen Stelle im Internet verfügbar. Um mit dieser schwierigen Informationslage umzugehen wird folgender Ansatz gewählt: Zunächst wird ein Überblick anhand zentraler Ergebnisse der 2010 abgeschlossenen Studie INNO-Appraisal dargestellt (Abschnitt 4.3.1). Darauf aufbauend wird die Kosten-Nutzen Analyse als eine Methode herausgegriffen, welche besonders deutliche Unterschiede in der Anwendung zwischen verschiedenen Innovationssystemen zeigt. Fallbeispiele aus den USA und Australien werden präsentiert (Abschnitt 4.3.2). Abschließend wird in überblicksartiger Form und in Anlehnung an die Präsentation österreichischer Evaluierungen in Abschnitt 4.3.3 eine selektive Zusammenschau aktueller, ökonomischer Evaluierungen aus anderen EU-Staaten gegeben.

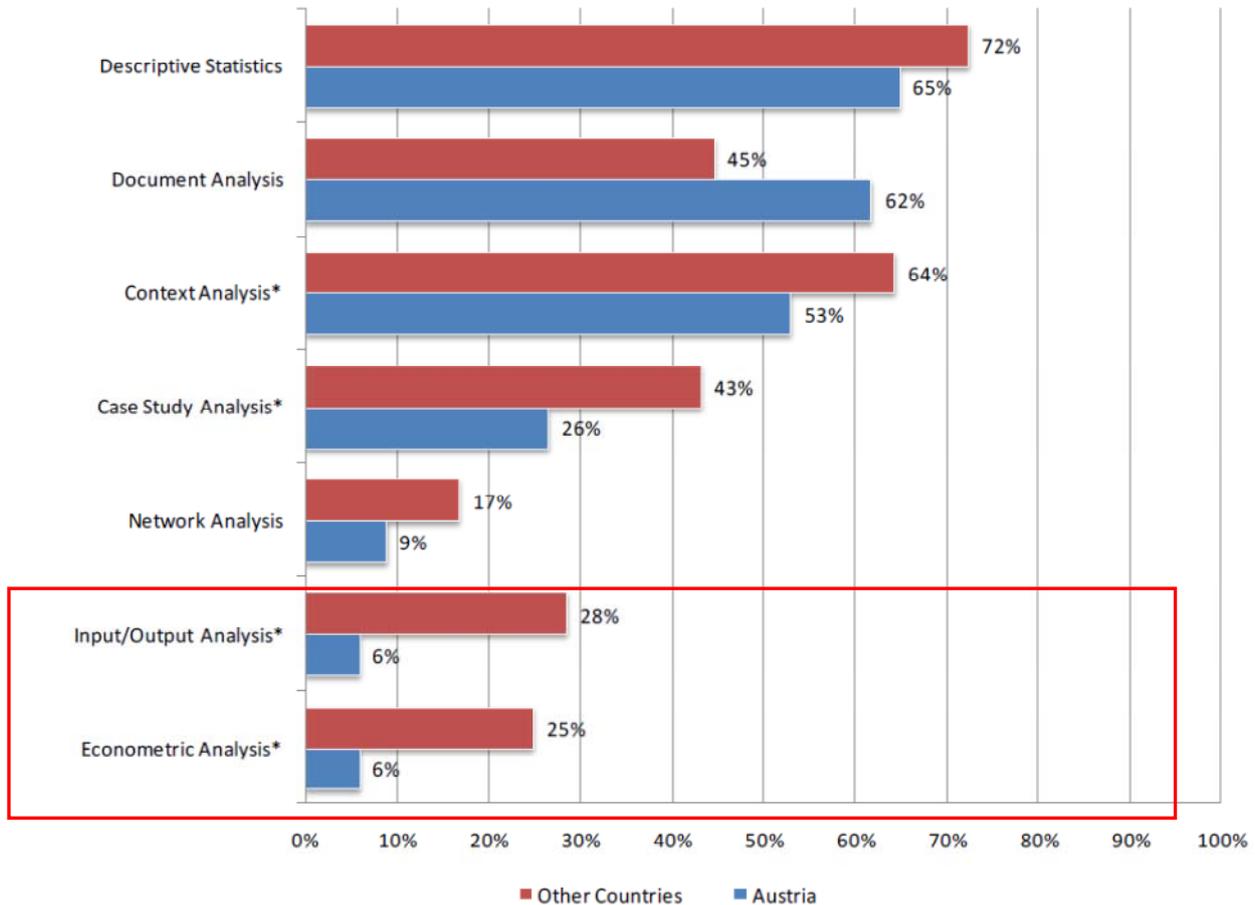
4.3.1 Österreichische FTI-Evaluation im internationalen Vergleich: Die Ergebnisse des INNO-Appraisal-Projekts

Abbildung 12 zeigt einen Vergleich der in Österreich verwendeten Methoden bei FTI-Evaluierungen mit den anderen EU-Staaten. Die mit einem Sternchen versehenen Kategorien zeigen dabei signifikante Unterschiede zwischen Österreich und den internationalen Vergleichsländern an. Etwas überraschen ist der signifikant geringere Einsatz von Fallstudien sowie von Kontextanalysen. Den Erwartungen entsprechend zeigt sich aber vor allem ein signifikant geringerer Einsatz ökonomischer Methoden sowie von Input-Output-Analysen in Österreich. Vor allem die wesentlich geringere Anwendung ökonomischer Verfahren aus Sicht der StudienautorInnen problematisch.

Diese Ergebnisse wurden bislang vor allem mit der Datenlage in Österreich erklärt. Warum aber existiert diese schlechte Datenlage? Eine Ursache hierfür könnte eine eher negative Einstellung gegenüber formalen, modellbasierten Methoden in Österreich sein. Hierfür gibt es anekdotische Evidenz. Eine quantitative, empirische Evidenz kann der Vergleich der Ergebnisse von Abbildung 12 und Abbildung 13 (siehe Abschnitt 3.1.2) liefern. Abbildung 12 konstatiert eine signifikant geringere Anwendung formaler, modellbasierter Evaluationsmethoden in Österreich, während

Abbildung 13 zeigt, dass dies keineswegs negativ wahrgenommen wird. Dieses Ergebnis konnte auch in den Interviews repliziert werden.

Abbildung 12: Angewandte Methoden bei FTI-Evaluierungen in Österreich und Vergleichsländern



INNO-Appraisal 2010, *Chi-Quadrat Statistik ist Signifikant auf dem 0.05 Niveau.

Eine ähnliche, letztlich soziale und kulturelle Strukturen reflektierende Feststellung trifft Schmidt (2007: 4) für die deutsche Situation, wenngleich in Deutschland ökonomische Methoden freilich wesentlich öfter angewandt und breiter akzeptiert scheinen als dies in Österreich der Fall ist: „Nonetheless, in some countries (e.g. the United States) policy makers and administrators tend to demand sophisticated, methodologically rigorous economic research as a basis for policy advice, while in other countries (e.g. Germany) they even struggle with researchers over side issues such as data confidentiality and the ethical basis of social experiments” (Schmidt 2007:4). Tatsächlich lassen sich in der methodischen Entwicklung relative deutliche Unterschiede zwischen den angelsächsischen Ländern USA und Großbritannien einerseits und den kontinentaleuropäischen Ländern andererseits feststellen. Ein cursorischer Überblick über die Herkunft ökonomischer Innovationen zeigt, dass so gut wie alle methodischen Innovationen aus den USA oder Großbritannien stammen. Die Frage ist, wie rasch bzw. ob überhaupt diese Methoden in andere Staaten diffundieren. Traditionell sind etwa die skandinavischen Staaten eng an den

angelsächsischen Diskursen orientiert; ähnliches gilt auch für die Schweiz und zunehmend für Deutschland. In Österreich, wie auch teilweise in Deutschland, spielt sicherlich aber auch die Ausbildung von SoziologInnen, PolitikwissenschaftlerInnen und VolkswirtInnen eine wichtige Rolle, deren Curricula traditionell eine eher schwache methodische Komponente aufwiesen. Somit könnte ein weiterer Grund für die seltene Anwendung formaler, modellbasierter Methoden auch eine Folge einer spezifischen wissenschaftshistorischen Orientierung in den relevanten Disziplinen sein. Nachdem hier aber die letzten Jahre durchaus entscheidende Änderungen in der Ausbildung brachten, könnte dieser Faktor in der Zukunft an Bedeutung verlieren.

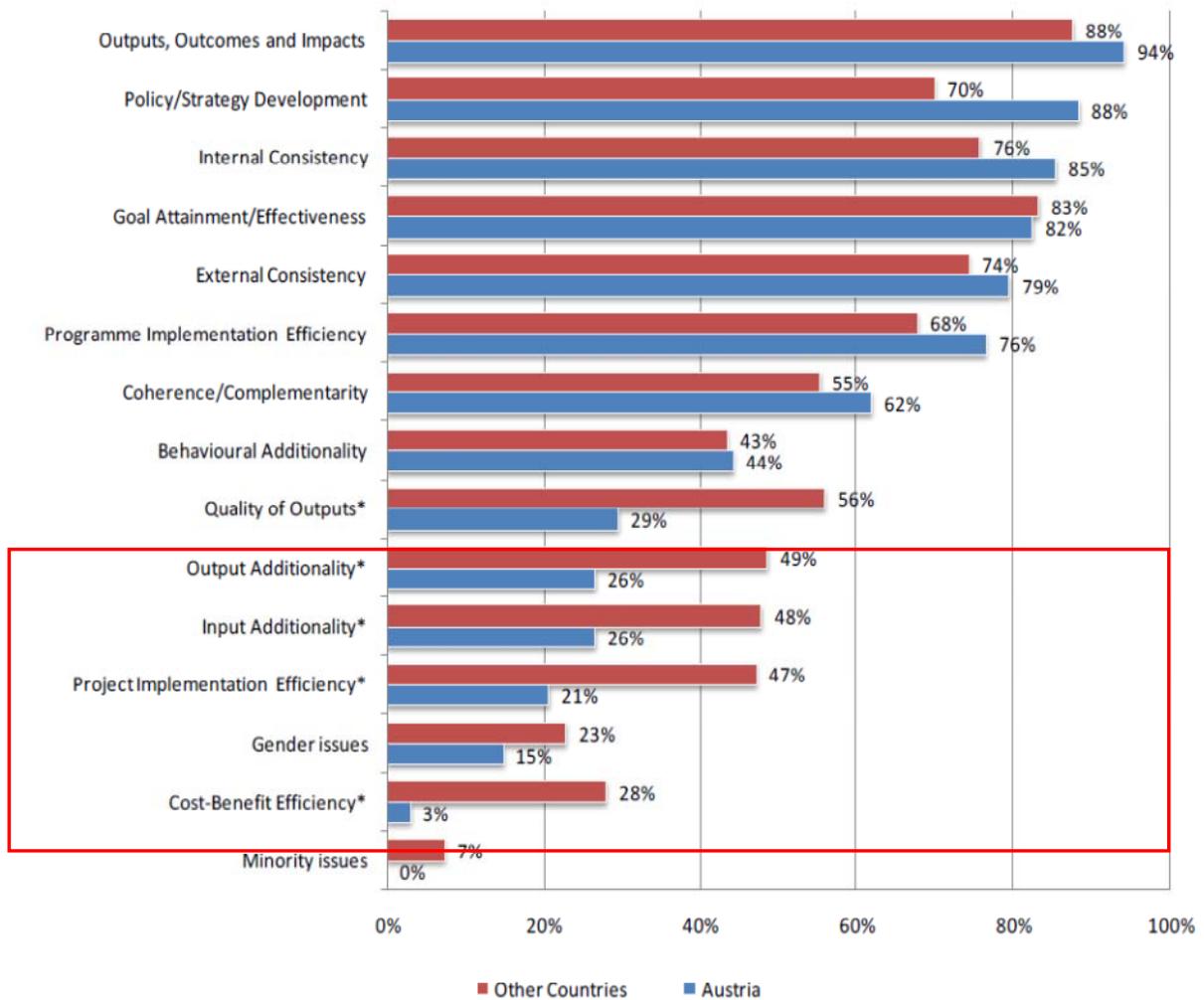
Die relativ seltene Anwendung ökonometrischer Methoden wird auch in den oben dargestellten Analysen zu den Evaluierungen der Jahre 2005-2010 in Österreich deutlich. Demnach gab es nur eine Studie mit einer ökonometrischen Anwendung (Streicher 2007). Dieser Befund wurde bereits in den Standards der Plattform Forschungs- und Technologieevaluierung im Jahr 2005 festgestellt: „Mikroökonometrische Arbeiten, die beispielsweise die Entwicklung von performance-Indikatoren von Firmen mit der Teilnahme an einem Technologieprogramm in Verbindung setzen gibt es in Österreich noch kaum.“ (fteval 2005:19). Auch Rhomberg et al. (2006) kommen zum gleichen Ergebnis: „Aus der methodischen Perspektive ist Österreich im internationalen Vergleich bei ökonometrischen Ansätzen vielfach noch ein Nachzügler.“ Insofern besteht diese methodische Schiefelage zuungunsten quantitativ-analytischer Methoden bereits seit längerem. Geändert hat sich daran in den letzten Jahren jedoch nichts, eher kam es zu einer weiteren Akzentuierung dieser Asymmetrie in der Methodenanwendung. In den Interviews wurde dieser Trend unter anderem mit einer zunehmenden Bedeutung von Interimsevaluierungen gegenüber ex-post Evaluierungen begründet. Es stellt sich jedoch die Frage, welche sachlogische Begründung hinter dieser Argumentation steht. Entscheidend dürfte wohl der Aspekt eines time-lags zwischen Intervention und Wirkungseintritt sein. Aber dieses Problem trifft für qualitativen Methoden in gleicher Weise zu. Ein weiteres Argument könnte auch das stärkere Gewicht auf formativer im Gegensatz zu summativer Evaluation bei Interimsevaluierungen sein.

Hier wäre jedenfalls kritisch anzumerken, dass eine methodisch rigorose Analyse der Effektivität eines Programms nicht nur am Ende eines Programms sinnvoll ist, insbesondere wenn ein Programm eine relativ lange Laufzeit aufweist.

Es mag auch fragwürdig erscheinen, Programme durch Mikromanagement ständig zu optimieren, wenn keine „harte“ Evidenz für die letztlich entscheidende Effektivität des Programms besteht.

Das Ergebnis von Abbildung 12 wird auch in den in Abbildung 13 dargestellten inhaltlichen Schwerpunkten von Evaluierungen in Österreich wiedergespiegelt. Demnach werden jene Fragestellungen in Österreich signifikant weniger als in den anderen EU-Staaten analysiert, die typischerweise die Anwendung analytischer, quantitativer Methoden voraussetzen. Diese inhaltlichen Aspekte sind insbesondere die Input- und Output-Additionalität, aber auch Effizienz- und Kosten-Nutzen-Betrachtungen.

Abbildung 13: Inhaltliche Ausrichtung der Evaluationen



INNO-Appraisal 2010, *Chi-Quadrat Statistik ist Signifikant auf dem 0.05 Niveau.

4.3.2 Kosten-Nutzen Analyse als angelsächsisches Spezifikum

Während Effizienzbetrachtungen generell selten sind, zeigt sich in der Anwendung von Kosten-Nutzen-Analysen (KNA) ein signifikanter Unterschied zwischen den angelsächsischen und kontinentaleuropäischen Staaten.⁸ So wurde in Österreich nach unserem Wissen nach noch keine einzige KNA zur Evaluation von FTI-politischen Maßnahmen angewandt. Das gleiche gilt auch für Deutschland (INNO-Appraisal 2010). Demgegenüber werden in den USA einige wichtige innovationspolitische Programme standardmäßig mit KNA evaluiert. Prominente Beispiele in diesem Zusammenhang sind das Advanced Technology Program (ATP) sowie das das National Institute for Standards and Technology (NIST) (siehe Box 7).

⁸ Für eine aktuelle deutschsprachige Einführung zur KNA siehe Hanusch (2010). Eine hervorragende, mit zahlreichen Praxisbeispielen versehene englischsprachige Einführung bietet das Handbook of Cost-Benefit Analysis des australischen Finanzministeriums [<http://www.finance.gov.au/obpr/cost-benefit-analysis.html>, 20.10.2011]. Eine kurze Einführung zur KNA aus FTI-Perspektive findet sich auch in der RTD-Evaluation Toolbox.

Box 7: ATP und NIST als Beispiele für innovationspolitische Institutionen, die KNA als Evaluationsmethode anwenden

ATP war ein US-amerikanisches Technologieprogramm, welches vermutlich zu den am häufigsten und auch professionellsten evaluierten innovationspolitischen Maßnahmen weltweit zählt. Es wurde 1990 mit dem Ziel gegründet, die Lücke zwischen Forschung und Vermarktung zu schließen. „ATP awards are selected through open, peer-reviewed competitions. All industries and all fields of science and technology are eligible. Proposals are evaluated by one of several technology-specific boards that are staffed with experts in fields, such as biotechnology, photonics, chemistry, manufacturing, information technology, or materials. All proposals are assured an appropriate, technically competent review even if they involve a broad, multi-disciplinary mix of technologies.“ (<http://www.atp.nist.gov/>, 20.10.2011). Gefördert sollen vor allem jene Projekte, die eine möglichst große Differenz zwischen privatem und sozialem Nettonutzen aufweisen. Trotz aller Evaluierungen, die durchwegs Erfolge des Programms nachweisen konnten, beendete Präsident George Bush 2007 das ATP-Programm.

Das NIST wurde 1901 gegründet und ist eine Subeinheit des U.S. Departments of Commerce. Es ist vorwiegend mit der Bereitstellung und Förderung unterschiedlicher Formen technologischer Infrastruktur befasst. Dabei arbeitet es eng mit dem Unternehmenssektor zusammen und entwickelt Standards und Messmethoden. Von besonderer Bedeutung sind dabei sogenannte „Infratechnologies“. Dabei handelt es sich vorwiegend um Messmethoden oder Testmethoden, die zur effizienten Verknüpfung von arbeitsteilig organisierten Produktionssystemen von Relevanz sind (Link, Scott:2011).

<http://www.nist.gov/index.html>

Das ATP hat bis zum Jahr 2006 14 Kosten-Nutzen-Analysen über 32 ATP-Projekte durchgeführt (Powell 2006). Dabei handelt es sich ohne Ausnahme um eine Weiterentwicklung der traditionellen KNA, weil hier nicht ein Projekt, sondern eine Summe von Projekten zu einem Cluster fusioniert und gemeinsam einer KNA unterzogen wurde (Powell 2006). Alle Evaluierungen mittels KNA wurden als externe Evaluation durchgeführt. Einen Auszug aus den Kosten-Nutzen-Analysen des NIST zeigt Tabelle 11.

Nachdem die von NIST entwickelten Infra-Technologien (siehe Box 7) Auswirkungen auf ganze Produktionsketten haben, wird in den KNA des NIST versucht, die Effekte entlang einer solchen Produktionskette zu erfassen. Die zweite Spalte von zeigt die in der jeweiligen Evaluierung berücksichtigten Abschnitte der Produktionskette. Kritisch ist anzumerken, dass zwar jede Evaluierung eine zusammenfassende Maßzahl zur Bewertung des Projekts nennt (Spalte drei von Tabelle 12), jedoch eine Vergleichbarkeit dieser Maßzahlen in den Erörterungen zur Methode explizit als problematisch bezeichnet wird. Die Ursache hierfür liegt in der differierenden Datenverfügbarkeit zwischen den Projekten. Nachdem aber eine Monetarisierung der Kosten- und Nutzenkomponenten an die Verfügbarkeit entsprechend detaillierter Daten gebunden ist, sind die Unterschiede in den Benefit-to-Cost-Ratios nicht nur auf unterschiedliche Projektperformances sondern auch eine Folge unterschiedlicher Datenverfügbarkeiten zurückzuführen.

„For that reason it is crucial that each analysis continue to identify all effects qualitatively and state clearly what is missing from the estimated evaluation metrics and emphasize that readers should not simply compare benefit-to-cost ratios or other matrices across the analyses. Instead, the conservative metrics must be kept in mind when comparing the economic impacts reported in the analysis.“ (Link, Scott 2011: 7). Ein weiterer, aus evaluatorischer Perspektive kritischerer Punkt bezieht sich auf die Abschätzung der Nutzenkomponenten durch direkte Befragung der Programmteilnehmer. Dieser Aspekt wird in Kapitel 5 näher erörtert.

Tabelle 12: Kosten-Nutzen-Analysen von NIST-Projekten

Economic Impact Analysis	Stage of Economic Activity	Benefit-to-Cost Ratio*
“Economic Assessment of the NIST Thermocouple Calibration Program,” Planning Report 97-1	R&D, production, and commercialization; commercialization for users of thermocouples evaluated qualitatively only	3-to-1
“Economic Assessment of the NIST Alternative Refrigerants Research Program,” Planning Report 98-1	R&D, production, and commercialization; commercialization benefits, including the value of better quality products, largely unquantified	4-to-1
“Economic Assessment of the NIST Ceramic Phase Diagram Program,” Planning Report 98-3	R&D primarily, but also production (detecting and diagnosing abnormalities in production), and commercialization (because delays in the introduction of new materials were avoided); commercialization benefits of faster development of new products not quantified, and benefits for some segments of the ceramics industry not quantified	10-to-1
“Benefit Analysis of IGBT Power Device Simulation Modeling,” Planning Report 99-3	R&D, production, and commercialization; production benefits for applications manufacturers and commercialization benefits of product quality for end users evaluated qualitatively only	23-to-1

Link, Scott 2011

Eine weitere interessante Anwendung der KNA stammt aus Australien. Dort wird ein seit einigen Jahren eine innovationspolitische Initiative der australischen Weinindustrie regelmäßig durch hoch standardisierte Kosten-Nutzen-Analysen evaluiert. Die Grundlagen für die Analysen bieten zahlreiche Publikationen des australischen Finanzministeriums zur Durchführung von KNAs.⁹ Damit wird sowohl die Nachvollziehbarkeit ermöglicht als auch eine solide Basis für den Vergleich von Ergebnissen geschaffen.

Box 8 zeigt einige wichtige Elemente der KNA.

⁹ <http://www.finance.gov.au/obpr/cost-benefit-analysis.html>,

Box 8: Fallbeispiel KNA in Australien

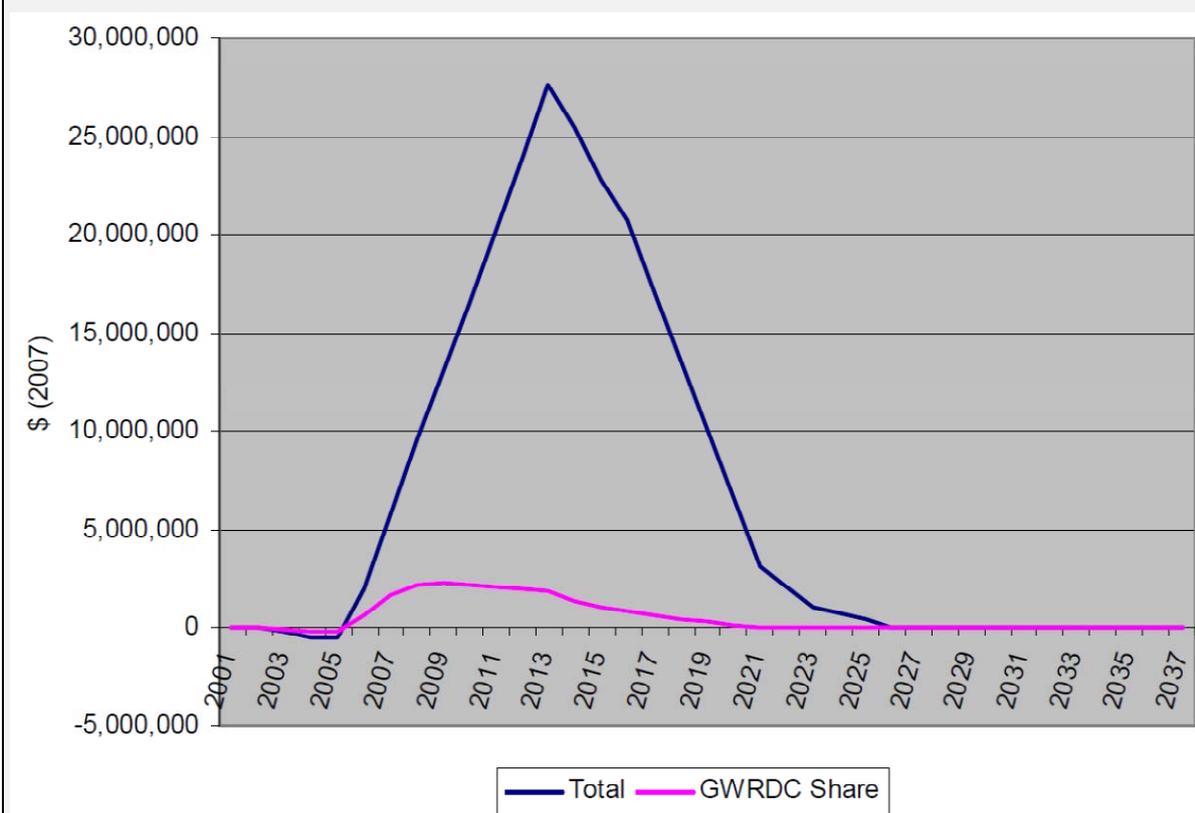
Die „Grape and Wine Research and Development Corporation“ (GWRDC) wurde 1992 als staatliche Institution gegründet. „The GWRDC supports the development of the Australian wine industry by planning and funding collective research and development programs and then facilitating the dissemination, adoption and commercialisation of the results throughout the industry. The GWRDC invests in R&D from existing providers (such as CSIRO, state agencies, universities and The Australian Wine Research Institute) on behalf of the Australian wine industry. The GWRDC's funding for R&D investment comes from levies on the annual grape harvest and wine yield, with the Australian Government matching these fund“ (<http://www.gwrdc.com.au/site/page.cfm?u=3>, 20.20.2011).

Zwischen 2002/03 und 2006/07 wurden insgesamt 40 sogenannte „project cluster“ durch GWRDC gefördert. Jeder Cluster besteht aus 1-13 sogenannten „component projects“. 2008 wurde dazu eine Evaluation von „Econsearch“ durchgeführt. Dazu wurden aus den 40 Clustern drei Cluster nach einem Zufallsprinzip ausgewählt und anhand einer standardisierten KNA evaluiert. Die folgenden Ausführungen beziehen sich auf einen der drei Cluster, den „vine physiology – flowering“ cluster. Dieser Cluster bestand aus 4 geförderten Projekten. Die Analyse beinhaltet einerseits eine KNA für den Cluster insgesamt als auch eine isolierte KNA für jedes der vier Projekte. Wichtige Merkmale der Analysen sind dabei folgende:

- „The CBA includes a base case and a counterfactual scenario, that is, the benchmark against which the ‚with GWRDC investment‘ scenario was compared. The base case was defined as what would have occurred without GWRDC investment in the technology or research.
- The CBA was conducted over a 30 year time period and results were expressed in terms of net benefits, that is, the incremental benefits and costs of the ‚with GWRDC investment‘ scenarios relative to those generated by the base case scenario.
- Costs and benefits were specified in real terms. Past and future values were converted to present values by applying a discount rate of 5%.
- In order to account for uncertainty, sensitivity analysis was undertaken using a range of values for key variables, including adoption profiles [of the invented technologies as part of the funded projects, d. Verf.] (Econsearch 2008).

Abbildung 14 zeigt die Ergebnisse der KNA nach Kosten und Nutzen mit einer Projektion bis 2037.

Abbildung 14: Nettonutzen der Forschungsergebnisse



Econsearch 2008

Um einen Eindruck von den dahinterstehenden Kosten- und Nutzenkomponenten zu bekommen, zeigt Abbildung 15 die identifizierten Komponenten für eines der insgesamt vier Teilprojekte des Clusters: „SAR 02/05 – Improving vineyard productivity through assessment of bud fruitfulness and bud necrosis.“ Die Daten, die zur Erfassung dieser Kategorien notwendig waren, stammen aus unterschiedlichen Quellen. Teilweise wurden sie von der Förderagentur aufbereitet, teilweise in qualitativen Interviews und Befragungen ermittelt. Interessant ist dabei insbesondere die Annahme über das kontrafaktische Szenario: Demnach nehmen die AutorInnen der Evaluation an, dass innerhalb von 10 Jahren die jeweils gleichen Ergebnisse wie bei staatlicher Intervention eintreten wären. Damit bringt die staatliche Intervention primär eine Beschleunigung des technologischen Wandels in der australischen Weinindustrie. Letztlich würden aber die Weinbauern die entsprechenden Innovationsaktivitäten auch ohne staatliche Förderung setzen.

Abbildung 15: Kosten- und Nutzenkomponenten von Projekt SAR 01/05

Scenario	Cost	Bearer of the Cost	Valued in Monetary Terms	Source of Information
Base case (without GWRDC investment) scenario	Identical to the 'with GWRDC investment' scenario but with a time lag of 10 years	See below	See below	See below
With GWRDC investment scenario	Project R&D costs	GWRDC and collaborators	Yes	GWRDC and collaborators
	Adoption costs (bud dissection analysis)	Winegrape growers	Yes	See text in section 4.3.1

Scenario	Benefit	Beneficiary	Valued in Monetary Terms	Source of Information
Base case (without GWRDC investment) scenario	Identical to the 'with GWRDC investment' scenario but with a time lag of 10 years	See below	See below	See below
With GWRDC investment scenario	Improved winegrape yield in varieties that are susceptible to PBN	Winegrape growers and winemakers	Yes	See text in section 4.3.2
	Other industry benefits (e.g. ensuring a more consistent supply of product)	Winegrape growers and winemakers	No	See text in section 4.3.2
	Spillover benefits (e.g. flow-on impacts from an improvement in the reliability of income levels at the vineyard level)	The broader community	No	See text in section 4.3.2

Econsearch 2008

Zusammenfassend lassen sich folgende Punkte zum Einsatz der KNA zur Evaluation innovationspolitischer Maßnahmen feststellen:

- Während in Kontinentaleuropa KNA gar nicht oder kaum als Evaluationsmethodik angewandt wird, ist deren Einsatz in den angelsächsischen Ländern häufig.
- Die Anwendungsgebiete umfassen vor allem größere Projekte in unterschiedlichen technologischen Bereichen wie etwa Luftfahrt, Raumfahrt, Medizintechnik, Infrastruktur, etc. Für Österreich würden sich demnach größere, klar abgegrenzte Projekte wie etwa Kompetenzzentren oder Raumfahrtprojekte als mögliche Anwendungsgebiete einer KNA anbieten.
- Traditionell wird die KNA im Vorfeld einer Projektdurchführung, etwa im Straßenbau, durchgeführt, um am Ende jenes Projekt zu selektieren, welches den höchsten

Nettonutzen aufweist. Bei den hier präsentierten FTI-Evaluierungen wird die KNA ausschließlich für ex-post Evaluierungen eingesetzt.

- Die Vorteile der KNA liegen sicherlich in der Tatsache, dass sie die Akteure zwingen, eine möglichst genaue Auflistung aller Nutzen und Kosten die mit einem Programm verbunden sind vorzunehmen. „ [Man] wird besser eine unvollkommene als keine Kosten-Nutzen-Analyse zu Rate ziehen, da sie immerhin einen Anlass zum Zusammentragen aller bedeutsamen Einflüsse, die von verschiedenen Alternativen zur Lösung eines Problems ausgehen, bietet und damit größere Transparenz verschafft sowie vor allem einen Argumentationszwang ausübt“ (Zimmermann, Henke 2005:104). Ein spezifisches Problem der Anwendung von KNA-Techniken im FTI-Bereich stellt die seriöse Berücksichtigung des sogenannten „Mit- und-Ohne Prinzips“ dar (Hanusch 2010). Demnach dürfen auf der Nutzenseite nur jene Faktoren berücksichtigt werden, die tatsächlich durch die politische Intervention verursacht worden sind. Dieses Prinzip ist etwa beim Bau öffentlicher Infrastruktur relativ einfach einzuhalten, weil große Projekte ohne staatliche Beteiligung mit höchster Wahrscheinlichkeit nicht realisiert werden würden. Dies ist im FTI-Sektor völlig anders. Hier ist die Frage, welches Ergebnis die Programmteilnehmer bei Nichtteilnahme realisiert hätten sehr komplex und Kernfrage jeder ernstzunehmenden Evaluation. Die hier KNAs aus den angelsächsischen Ländern befragen direkt die TeilnehmerInnen über die kontrafaktische Situation. Dies ist aber eine durchwegs problematische Strategie (siehe unten).
- Ein weiteres, spezifisches Problem von KNA-Analysen im FTI-Kontext besteht in der Berücksichtigung des time-lags zwischen Intervention und Programmwirkung. Während dieses Problem freilich für alle Evaluationsmethoden von Bedeutung ist, ist es bei der KNA deswegen besonders schwerwiegend, weil hier tatsächlich möglichst alle Nutzen berücksichtigt werden sollen und etwa nicht nur der Effekt auf den Umsatz eines programmteilnehmenden Unternehmens.

4.3.3 Überblick zu aktuellen ökonometrischen Evaluierungen innovationspolitischer Maßnahmen

Tabelle 13 zeigt eine Auswahl ökonometrischer Evaluierungen von innovationspolitischen Maßnahmen aus unterschiedlichen OECD Staaten von 2006-2011. Zunächst einmal kann festgestellt werden, dass es eine größere Anzahl solcher Arbeiten in anderen Staaten gibt und Österreich hier ein deutliches Defizit aufweist. So gab es etwa im besagten Zeitraum mit Streicher (2007) in Österreich nur eine einzige ökonometrische Arbeit, die noch ein Ausfluss der FFF-Evaluierung von 2004 war. Methodisch zeigen die von Tabelle 13 aufgelisteten Arbeiten eine deutliche Dominanz von Matching-Ansätzen. Weniger bedeutsam sind Instrumentenvariablenschätzer und Kontrollfunktionsansätze. Eine Darstellung der dafür notwendigen Datensets, ebenfalls mit internationalen Fallbeispielen, findet sich in Kapitel 7.

Analysiert man die Ergebnisse der ökonometrischen Studien in Tabelle 13, so ergibt sich in Summe ein optimistisches Bild der Fördereffekte. Lediglich 2 von insgesamt 19 Studien identifizieren mangelnde Wirkungen der Förderungen. Damit scheint die Wirksamkeit von F&E-Subventionen jedenfalls durch internationale Studien einigermaßen gut abgesichert.

Tabelle 13 Auswahl rezenter ökonomischer, internationaler Evaluationen von innovationspolitischen Interventionen

AutorInnen	Jahr	Journal/ Working Paper Series, #	Staat, Agentur , Förderinstrument	Methode	Ergebnis
Aerts/ Czarnitzki	2006	IWT-Studies, 54	Flandern, iwt	Matching, Instrumentenvariablenansatz	Geförderte Unternehmen geben das Geld in vollem Umfang für F&E aus (kein crowding out).
Cornet et al.	2006	CPB Discussion Paper, 58	Niederlande, Innovation voucher für KMUs	Randomisierte Allokation der Vouchers, Lineares Wahrscheinlichkeitsmodell	8 von 10 vouchers werden für Forschungsaufträge genutzt, die ohne Förderung nicht nachgefragt worden wären
Berube/ Mohnen	2007	UNU-MERIT Working Paper Series, 15	Kanada, Direkte und steuerliche F&E-Förderung	Matching	Unternehmen, die sowohl direkte und steuerliche Förderung erhalten, sind innovativer als Unternehmen, die nur steuerliche Förderung erhalten.
Clausen	2007	TIK Working Paper Series on Innovation Studies, 20070615	Norwegen, 5 Förderinstrumente werden untersucht, die eher marktnahe oder marktferne F&E-Projekte fördern	Instrumentenvariablenansatz	„Forschungssubventionen“ haben einen positiven Effekt auf unternehmerische F&E-Ausgaben, „Entwicklungssubventionen“ substituieren private Ausgaben.
Czarnitzki et al.	2007	Journal of Applied Econometrics	Deutschland und Finnland im Vergleich, Effekte von F&E-Subventionen (ja/nein) und Forschungsk Kooperationen	Matching	F&E-Kooperationen haben in beiden Staaten einen positiven Effekt auf Patentaktivitäten und F&E-Intensität. In Deutschland haben Subventionen für isolierte Forschungsprojekte einzelner Unternehmen keinen Effekt; für Finnland lässt sich demgegenüber ein positiver Effekt nachweisen.
Görg/ Strobl	2007	Economica	Irland, gesamtes staatliches F&E-Fördersystem	Matching, Differenz-von-Differenzen-Ansatz	Bei heimischen Unternehmen verursachen kleine Subventionen eine Zunahme der F&E-ausgaben, während große Subventionen ein crowding-out zur Folge haben. Bei ausländischen Unternehmen kann unabhängig vom Subventionsvolumen weder eine Zunahmen noch ein crowding-out identifiziert werden.

AutorInnen	Jahr	Journal/ Working Paper Series, #	Staat, Agentur, Förderinstrument	Methode	Ergebnis
Hussinger	2008	Journal of Applied Econometrics	Deutschland, gesamte bundesstaatliche F&E-Projektförderung	Kontrollfunktionsansatz	F&E-Subventionen haben einen positiven Effekt auf private F&E-Ausgaben pro Beschäftigten und auf den Umsatz innovativer Produkte.
Takalo et al.	2008	Bank of Finland Research Discussion Papers, 7	Finnland, direkte F&E-Förderung für Unternehmen durch Tekes	Vollständig spezifiziertes strukturelles Modell	Große Heterogenität der Maßnahmeneffekte; der return on subsidies für Tekes liegt bei 30-50%.
Wolff/ Reinthaler	2008	Research Policy	Panel von OECD-Staaten, F&E-Subventionen j/n	Instrumentenvariablenansatz	F&E Subventionen haben einen positiven Effekt auf F&E-Ausgaben, wobei diese um 20% stärker zunehmen als F&E-Beschäftigung.
Gonzalez/ Pazo	2008	Research Policy	Spanien	Matching	Absenz eines totalen oder partiellen crowding-out. Vor allem kleine und niedrigtechnologische Unternehmen hätten ohne F&E-Subvention gar keine F&E-Aktivitäten aufgenommen.
Cincera et al.	2009	European Economy, Economic Papers, 376	Makroökonomische Effizienzanalyse von EU- und OECD-Staaten	Stochastic-Frontier Analyse, Data Envelope Analyse	Staaten mit der besten Innovationsperformance weisen auch die höchste Effizienz bei F&E-Subventionen auf, während höhere staatliche F&E-Ausgaben in % des Konsums mit einer Effizienzreduktion assoziiert sind. Es gibt kein allgemein gültiges Modell zur Erhöhung der Effizienz staatlicher F&E-Förderung.
Clarysse/ Knockaert	2009	Iwt Studies, 64	Flandern, KMU-Förderungen	Kontrollfunktionsansatz	Die Faktoren zur Erklärung, ob ein Unternehmen einen Förderantrag stellt und jene, die entscheiden ob es eine Förderung bekommt, sind unterschiedlich.
Einiö	2009	Helsinki Center for Economic Research, Discussion Paper, 263	Finnland, alle Förderungen von Tekes für Unternehmen	Instrumentenvariablenansatz	Ein Euro öffentlicher Förderung induziert F&E-Mehrausgaben der Unternehmen in der Höhe von 1,5 Euro.

AutorInnen	Jahr	Journal/ Working Paper Series, #	Staat, Agentur, Förderinstrument	Methode	Ergebnis
Czarnitzki/ Bento	2010	ZEW Discussion Paper, 10-073	Belgien, Deutschland, Luxemburg, Südafrika, Spanien; Innovationsförderung j/n (CIS-Indikator)	Matching	Förderungen haben einen positiven Effekt auf die Innovationsaktivitäten der Unternehmen. Vier der fünf Staaten würden von einer weiteren Expansion ihres Fördersystems profitieren. Es zeigen sich keine Hinweise für eine systematische Fehlallokation der Förderressourcen.
Marino/Partta	2010	DRUID Paper	Dänemark, Förderung j/n, (CIS-Daten)	Matching mit kontinuierlicher Fördervariablen	Förderungen wirken positiv auf Input-Output- und Verhaltensadditionalität auf einem breiten Förderspektrum
Arvanitis et al.	2010	Science and Public Policy	Schweiz, Förderung kooperativer Forschungsprojekte durch die CTI	Matching	Förderungen haben einen positiven Effekt der umso größer ist, je höher die Förderung ausfällt
Canter/ Kösters	2011	Small Business Economics	Deutschland/Thüringen, Förderungen für start-ups	Logistische Regression	Förderagenturen folgen einer „picking the winner“-Strategie; trotzdem zeigen sich bei Aggregation aller Förderungen auch Probleme mit dieser Strategie, was die Grenzen eines „precise policy targeting“ aufzeigt
Eckl/ Engel	2011	Journal of Technology Transfer	Deutschland, Programm „Industrielle Gemeinschaftsforschung“	Multinomiales Probit Modell, Instrumentenvariablenansatz	Unternehmen mit engen Kontakten zu WissenschaftlerInnen, die an Projekten industrieller Gemeinschaftsforschung arbeiten, profitieren von knowledge spillovers auch wenn sie nicht an dem Projekt offiziell teilnehmen
Czarnitzki et al.	2011	Research Policy	Kanada, steuerliche F&E-Förderung	Matching	Steuerliche Förderung induziert zusätzliche Innovationsaktivitäten

5 Evaluierungsmethoden im FTI-Kontext: Theorie, internationale Fallbeispiele und Anwendung in Österreich

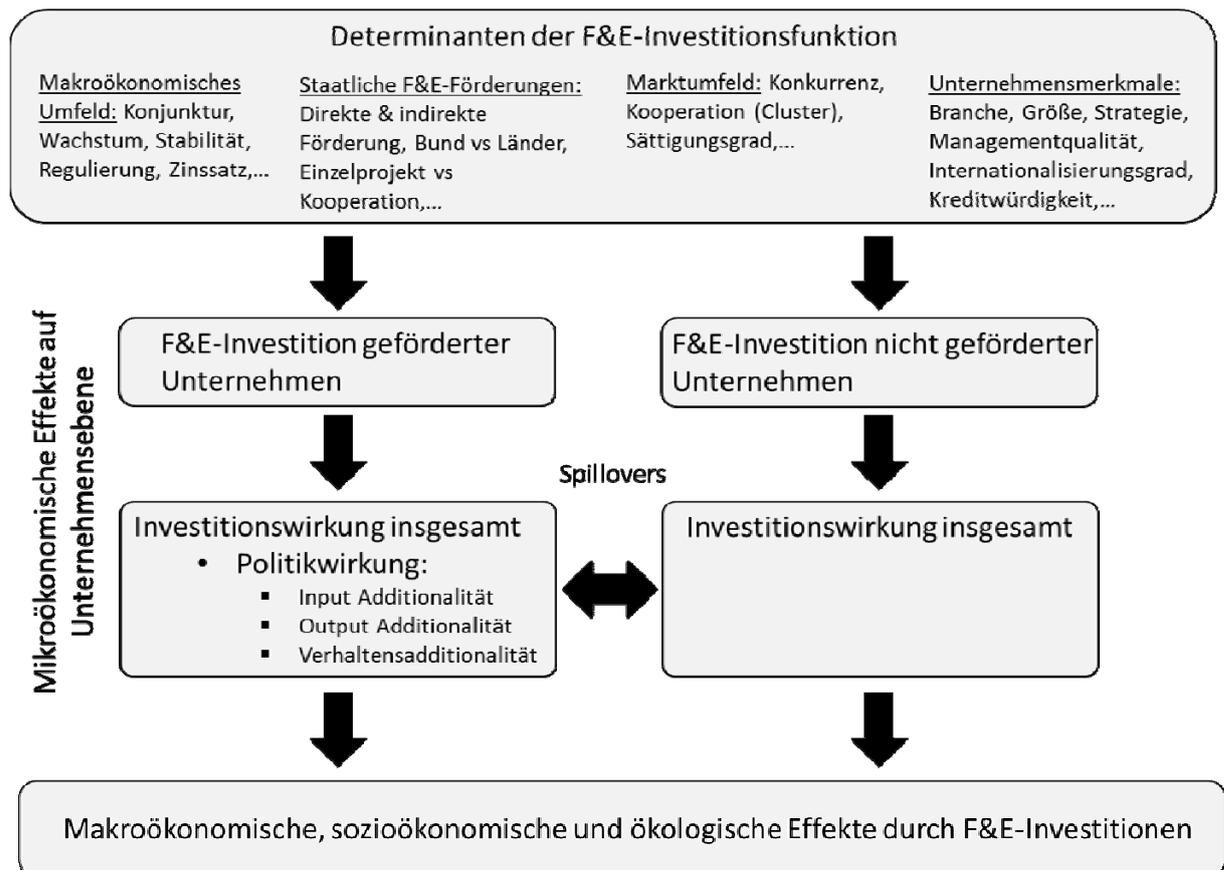
„It is tempting for the policymaker [and for the funding agency, d. Verf.] to compile dossiers replete with ‘success stories’ which may be used to justify continuation or expansion of the policy measure in question.“

Georghiou (2004:58)

Die Evaluation von innovationspolitischen Maßnahmen macht den Einsatz unterschiedlicher Methoden notwendig. Der Anspruch nach nachvollziehbaren und robusten Ergebnissen verlangt eine Reflexion über den Methodeneinsatz sowie die ständige Suche nach neueren, überlegeneren Methoden. Dies ist nicht zuletzt auch eine Folge der komplexen Zusammenhänge, die Evaluierungen berücksichtigen müssen. Die einfache Frage nach den kausalen Effekten politischer Intervention ist gerade im FTI-Bereich mit zahlreichen Problemen verbunden. Eine methodisch unreflektierte Antwort auf diese Frage wird systematisch falsche Ergebnisse erbringen.

Einen Eindruck von der komplexen Gemengelage FTI-politischer Intervention gibt Abbildung 16. Diese fasst zunächst im oberen Teil die wichtigsten Determinanten der F&E-Investitionsfunktion zusammen. Betrachtet man diese sicherlich nicht vollständige Aufzählung, so wird deutlich, dass staatliche Förderung nur ein Faktor neben vielen anderen ist, d.h. die Frage ob, wie viel und in welches Technologiefeld ein Unternehmen investiert, ist nur in einem bestimmten Ausmaß auf politische Intervention zurückzuführen. Entsprechend investieren freilich nicht nur geförderte Unternehmen in F&E, sondern auch nicht-geförderte Unternehmen. Schließlich ist eine F&E-Investition auf vielen Märkten mittlerweile *conditio sine qua non* zur Erlangung von Wettbewerbsvorteilen. Bei den mikroökonomischen Effekten auf Unternehmensebene ist bei den geförderten Unternehmen zunächst einmal zu differenzieren in die Effekte der Investition insgesamt und die Politikeffekte. Es wäre ein großer Fehler, den ganzen Effekt einer F&E-Investition auf die politische Förderung zurückzuführen. Tatsächlich ist es eine der Hauptaufgaben der Evaluation, diese beiden Effekte auseinanderzuhalten. Generell gesprochen ist der Politikeffekt eine Teilmenge des F&E-Investitionseffekts. Von Bedeutung ist auch die unterschiedliche Wahrnehmung von Politik und Unternehmen über die Funktion eines geförderten Projekts. Während die förderpolitische Vorstellung darauf basiert, dass es sich hierbei um ein klar abgegrenztes Projekt mit einem bestimmten Input und Output handelt, stellt sich die Perspektive aus Unternehmenssicht häufig völlig anders dar. Demnach ist das geförderte Projekt lediglich Teil einer längerfristigen Unternehmensstrategie, die eine Summe von Leistungen bündelt, wobei das geförderte Projekt nur eines von mehreren, eng zusammenhängenden unternehmerischen Initiativen darstellt. Bislang wurde jedoch in der Evaluationsforschung noch keine einheitliche methodische Antwort auf diesen als „Project Fallacy“ bezeichneten Sachverhalt gefunden (Georghiou 2007). Eine intime Kenntnis unternehmerischer Strategien ist aus mehreren Gründen nur schwer zu erlangen.

Abbildung 16: Determinanten des F&E-Investitionsverhalten ab und dessen Auswirkungen?

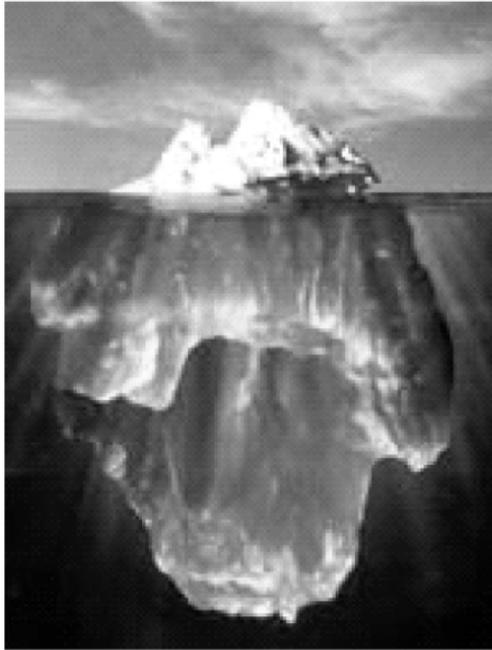


Eigene Darstellung

Der kausale Fördereffekt kann im geförderten Unternehmen wiederum auf der Input- oder der Output Seite erfasst werden. Die methodische Forschung der letzten Jahre hat insbesondere auch verhaltensbezogene Auswirkungen als relevante Fördereffekte thematisiert (Falk 2006). Zwischen den geförderten und nicht geförderten Unternehmen kommt es zu Spillovers, deren Erfassung ebenfalls große methodische Herausforderungen bereitet (Klette et al. 2000). Dabei ist zu berücksichtigen, dass die Spillovers nicht nur in Form positiver Wissensspillover auftreten, sondern auch negative Effekte haben können, wenn etwa geförderte Unternehmen nicht-geförderte Unternehmen vom Markt verdrängen, weil sie einen Kostenvorteil aufweisen. Ausgehend von den mikroökonomischen Effekten folgen weitere Effekte auf makroökonomischer Ebene sowie im ökologischen und sozialen Subsystem einer Gesellschaft. Damit macht Abbildung 16 deutlich, dass eine Evaluation einerseits unterschiedliche Effekte berücksichtigen muss und andererseits eine Ermittlung des politischen „Nettoeffekts“ keine triviale Aufgabe darstellt. Bezogen auf die Variablen, welche die unterschiedlichen Wirkungen von Förderungen erfassen, ist festzustellen, dass nur ein Teil davon unmittelbar beobachtet werden kann. Ein Großteil der Wirkungen bleibt daher typischerweise außerhalb traditioneller FTI-politischer Evaluation. Dieses Problem wird im sogenannten „Eisbergmodell“ in Abbildung 17 dargestellt. So wird hier etwa deutlich, dass neben Spillovers auch die wichtige Frage nach dem Nutzen von Innovationen für die KonsumentInnen

meist nicht beachtet wird. Freilich, im ökonomischen Paradigma ist diese Komponente nirgends besser abgebildet als durch den Umsatz der betreffenden Innovation.

Abbildung 17: Das Eisbergmodell: Nur ein Teil der Fördereffekte lässt sich "einfach" beobachten



Sales of innovative product
Reduced process costs
Licence income

Use of technology in other parts of the business
New contacts/networks & prestige
Organisation and method learning
Competence & training

Spillovers to non-participants
User and social benefits

Georgiou 2007

Angesichts dieser komplexen Sachlage hat sich bei FTI-Evaluierungen ein umfangreiches Methodenset etabliert, welches in Abbildung 18 überblicksartig dargestellt ist. Insbesondere hat sich im Einklang mit der methodischen Forschung in anderen sozialwissenschaftlichen Disziplinen die Erkenntnis durchgesetzt, dass quantitative und qualitative Methoden einander in sinnvoller Art und Weise ergänzen.

Abbildung 18: FTI-Evaluationsmethoden im Überblick

<p>Statistical data analysis:</p> <ul style="list-style-type: none"> • Innovation Surveys: provides basic data to describe the innovation process, summarised using descriptive statistics. • Benchmarking/Ranking: allows performing comparisons based on a relevant set of indicators across entities providing a reasoned explanation their values.
<p>Modelling methodologies:</p> <ul style="list-style-type: none"> • Macroeconomic modelling and simulation approaches: allows estimating the broader socio-economic impact of policy interventions. • Microeconomic modelling: permits to study the effect of policy intervention at the level of individuals or firms. There are mechanisms to control for the counterfactual by specifying a model which allows estimating the effects on the outcome of the participant had the programme not taken place. • Productivity analysis: permits to assess the impact of R&D on productivity growth at different levels data aggregation. This is particularly relevant to analyze the broader effects of R&D on the economy. • Control group approaches: allows capturing the effect of the programme on participants using statistical sophisticated techniques.
<p>Qualitative and semi-quantitative methodologies:</p> <ul style="list-style-type: none"> • Interviews and case studies: uses direct observation of naturally occurring events to investigate behaviours in their indigenous social setting. • Cost-benefit analysis: allows establishing whether a programme or project is economically efficient by appraising all its economic and social effects. • Expert Panels/Peer Review: measures scientific output relying on the perception scientists have of the scientific contributions made by other peers. Peer review is the most widely used method for the evaluation of the output of scientific research. • Network Analysis: allows analyzing the structure of cooperation relationships and the consequences for individuals' decisions' on actions providing explanations for the observed behaviors by analyzing their social connections into networks. • Foresight/ Technology Assessment: used to identify potential mismatches in the strategic efficiency of projects and programmes.

RTD Evaluation Toolbox 2002

Angesichts der Methodenfülle konzentriert sich diese Studie auf einige wenige ausgewählte Methoden und stellt aktuelle methodische Weiterentwicklungen vor. In Abschnitt 5.1 werden qualitative Methoden mit Schwerpunkt qualitative Netzwerkanalyse und Fokusgruppen diskutiert. Im darauf folgenden Unterkapitel 5.2 wird das mikroökonomische Paradigma der Programmevaluation präsentiert. Abschnitt 5.3 stellt schließlich noch die Frage, welche Methoden für welche Form von politischer Intervention besonders gut geeignet erscheinen.

5.1. QUALITATIVE METHODEN IN FTI-WIRKUNGSANALYSEN

Die letzten zehn Jahre zeigten eine verstärkte Integration qualitativer Ansätze in Evaluationen allgemein, aber auch spezifisch in Wirkungsanalysen FTI-politischer Programme. Diese gewinnen aufgrund ihrer spezifischen Vorteile gegenüber formalen, strukturierten Verfahren sowohl als Ergänzung oder Kombination zu quantitativen Verfahren (Stichwort: „mixed methods“, „triangulation“), aber auch als eigenständige Methoden immer mehr an Bedeutung.

Im Vergleich zur quantitativen Evaluationsforschung zeichnet sich die qualitative Evaluationsforschung durch eine größere Offenheit und Flexibilität aus. Ihre Vorgehensweise ist frei, explorativ und hypothesengenerierend. Das Verstehen von Veränderung sowie die Erklärung von subjektiven Verhaltensweisen stehen dabei im Vordergrund. Qualitative Verfahren weisen eine hohe Inhaltsvalidität sowie einen tiefen Informationsgehalt auf.

„Sie sind insbesondere dann wichtig, wenn keine klaren Effizienzkriterien aufzustellen sind und wenn die Veränderungen in einem komplexen, sich verändernden Praxisfeld stattfinden“ (Mayring, 2002:63).

Die qualitative Evaluationsforschung sieht es vor, „Praxisveränderungen wissenschaftlich zu begleiten und auf ihre Wirkungen hin einzuschätzen, indem die ablaufenden Praxisprozesse offen, einzelfallintensiv und subjektorientiert beschrieben werden“ (ibid.)

Box 9: Kennzeichen qualitativer Evaluationsforschung nach Mayring (2002)

Dabei wird auf folgende Aspekte besonders Wert gelegt:

1. In qualitative Evaluationen soll es auch möglich sein, aus den beobachteten Prozessen heraus neue Bewertungskriterien induktiv aufzustellen und zu begründen
2. Die in der Praxis beteiligten Personen sollten selbst zu Wort kommen und an der Evaluation beteiligt werden
3. Eine offene, holistische Schlussbewertung am Ende einer Evaluation analysiert einerseits die Wirkungen z.B. eines Programms in Bezug auf die vorab gesetzten Ziele, fasst aber andererseits auch den Gesamteindruck der abgelaufenen Veränderungen auf eine ganzheitliche Art und Weise zusammen (Inklusion im Evaluationsprozess neu entstandener Bewertungskriterien).

Wann können qualitative Methoden in FTI-Evaluationen eingesetzt werden? Qualitative Verfahren werden in allen drei Phasen einer Projektevaluation eingesetzt:

- **Ex-ante:** mittels qualitativer Methoden wird das Design einer Wirkungsanalyse erstellt. Die Identifizierung von relevanten Fragestellungen und das Entwickeln von Hypothesen stehen im Vordergrund. Ex-ante wird auch analysiert, wie und warum ein Programm auf das jeweilige Zielsystem wirken kann.
- **Interim:** In diesem Stadium dienen qualitative Methoden dazu, Einblicke in den aktuellen Stand des Programms und in erste Zwischenergebnisse rasch und detailliert zu bekommen.
- **Ex-post:** In dieser Phase ermöglichen qualitative Verfahren die Erklärung von Wirkungszusammenhängen und Kontexten, die für die Wirkungsabschätzung besonders bedeutend sind. Weiters können komplexe, unterschiedliche Effekte bei unterschiedlichen TeilnehmerInnen/Zielsystemen beschrieben und analysiert werden.

Die Schwächen qualitativer Verfahren sind, dass diese sowohl zeit- als auch kostenintensiv sind und hohe methodische Kompetenz bei der Erhebung und Auswertung der Daten voraussetzen. Weiters gilt es auch zu beachten, dass aus den qualitativen Daten keine statistisch repräsentativen Ergebnisse erzielt sowie keine zahlenmäßigen Mengenangaben gemacht werden können.

Aufgrund dieser Nachteile qualitativer Verfahren ist häufig eine Kombination quantitativer und qualitativer Methoden sinnvoll, um programmspezifische Wirkungen umfassend und exakt zu messen und die Güte und Qualität der Ergebnisse zu sichern.

5.1.1 Gemeinsamkeiten qualitativer Methoden

In Wirkungsanalysen kommen verschiedene qualitative Methoden zum Einsatz. Qualitative Interviews (problemzentriertes Interview, narratives Interview, etc.) sowie Gruppendiskussionen (ExpertInnen-Interviews, Fokusgruppen, etc.) werden sehr häufig verwendet.

Ebenso stellen Fallstudien (historische und deskriptive Fallstudien) ein Standardelement in Evaluationen dar. Fallstudien bezeichnen nicht eine qualitative Methode per se, sondern beschreiben den Einsatz unterschiedlicher Erhebungsmethoden (Interviews, Beobachtung, Desk Research,...), die einen spezifischen Kontext aus verschiedenen Perspektiven zu beleuchten.

Die qualitative Methode der Beobachtung (verdeckt, offen, teilnehmend, nicht-teilnehmend) wird eher selten verwendet. Nach Hollstein (2007) besteht der gemeinsame Nenner aller qualitativer Verfahren in den folgenden vier Bedingungen:

- Soziale Realität ist immer konstruiert. Dies bedeutet, soziale Realität wird nicht als „vorgegeben“ betrachtet, sondern im jeweiligen Situationskontext entwickelt (s. Thomas Theorem)
- Soziale Realität ist sinnhaft geordnet. Personen handeln in Bezug auf einen Sinn, nach dem sie ihr Handeln ausgerichtet haben (Symbolischer Interaktionismus nach G. Mead und H. Blumer).
- Soziale Realität ist immer räumlich und zeitlich gebunden. Der jeweilige Kontext ist entscheidend.
- Soziale Realität ist prozesshaft. Soziale Realität wird immer ausgehandelt.

Das Hauptziel der verschiedenen qualitativen Verfahren ist das „Sinnverstehen“. Darunter versteht man „das methodisch kontrollierte Fremdverstehen“. Der Einsatz sowohl offener Erhebungsverfahren, mit denen die individuelle Relevanz und der Bedeutungskontext erfasst werden kann als auch die Anwendung einer interpretativen Auswertungslogik sind Voraussetzung dafür, dieses Hauptziel (Sinnverstehen) zu erreichen (siehe Hollstein 2007).

In der Folge werden aus dem breiten Spektrum qualitativer Methoden in der Evaluationsforschung zwei Verfahren näher vorgestellt: die qualitative soziale Netzwerkanalyse und innovative Fokusgruppen.

5.1.2 Die qualitative soziale Netzwerkanalyse

Netzwerkanalysen haben sich in den letzten Jahren in den verschiedensten wissenschaftlichen Disziplinen etabliert (Hollstein/Straus, 2006). Auch als Analyse-Tool in Evaluationen kommen sie international regelmäßig zum Einsatz. In österreichischen Wirkungsanalysen FTI-politischer Programme werden vorwiegend quantitative Verfahren der Netzwerkforschung verwendet, qualitative soziale Netzwerkanalysen oder die Kombination aus quantitativen und qualitativen Elementen haben dieses Forschungsfeld noch nicht erobert. Nachfolgend wird die Methode der qualitativen sozialen Netzwerkanalyse näher vorgestellt, ihre Stärken und Vorgehensweise, um das Potential dieses Verfahrens auch für den Bereich der FTI-Evaluationen sichtbar zu machen.

Die quantitative Netzwerkanalyse und ihre Grenzen

Formale (quantitative) Netzwerkverfahren analysieren die Struktur (z.B. Dichte) von Netzwerken, die Leistung der Akteure und die Relationen zwischen den verschiedenen Beziehungen in einem Netzwerk. Diese arbeiten mit hochstandardisierten Erhebungsverfahren und elaborierten, komplexen Modellrechnungen in der Analyse (ibid.: 12).

Aufgrund des hohen Standardisierungsgrades liegt die Stärke formaler Netzwerkanalysen vor allen in der Prüfung konkreter Hypothesen und der Beantwortung eng zugeschnittener Fragestellungen. Ein hoher Standardisierungsgrad bringt aber aus forschungspraktischen Gründen folgende Einschränkungen mit sich:

1. Festlegung der Art der Netzwerkakteure: die Art der Netzwerkakteure, z.B. ob Verwandte, ArbeitskollegInnen oder Vorgesetzte analysiert werden, muss exakt definiert werden. Oft muss in diesem Zusammenhang auch eine Einschränkung in Bezug auf die Anzahl der Personen vorgenommen werden sowie in Bezug auf den Kontakt-Level zwischen den AkteurInnen (ibid.: 16).
2. Festlegung der Beziehungsinhalte: auch eine klare Definition der Beziehungsinhalte (z.B. materielle Unterstützung, berufliche Kooperation etc.) muss vorab bereits festgelegt sein.

Die Entscheidung für diese Einschränkungen bedeutet „die Entscheidung für ein eingeschränktes Handlungs- und Akteursmodell, welches zum Beispiel unterschiedlichen Wahrnehmungen, Interessen und Orientierungen der AkteurInnen nur eingeschränkt Rechnung trägt“ (Hollstein, 2007: 3361). Bestimmte Potentiale des Netzwerkansatzes bleiben so ungenutzt.

Eine rein formale Netzwerkanalyse vernachlässigt daher die Subjektperspektive und kann die Art und Weise des Aufbaus von Kooperationen oder handlungslimitierende und -ermöglichende Kontextfaktoren nicht erklären (ibid.)

Bezogen auf diese Aspekte kann die qualitative Sozialforschung einen wertvollen Beitrag zur Netzwerkforschung leisten.

Der Ertrag qualitativer Verfahren für die Netzwerkforschung

Qualitative Verfahren der Netzwerkforschung nähern sich der sozialen Realität/ dem Forschungsgegenstand offen. Durch das Setzen eines Stimulus (z.B. Wem fühlen Sie sich emotional verbunden?) wird versucht, sich schrittweise dem Relevanzsystem des/der Befragten anzunähern ohne dass Festlegungen vorab getroffen oder Beziehungsinhalte definiert werden (wie z.B. bei der Frage: Wer gehört zur Familie? Wen fragen Sie um Rat?) (vgl. Hollstein/Straus, 2006).

Die Akteursperspektive bildet den Mittelpunkt des Forschungsinteresses. Qualitative Verfahren können daher zahlreiche wertvolle Beiträge zur Netzwerkforschung liefern (ibid.; Hollstein, 2007):

Box 10: Wichtige Beiträge qualitativer Verfahren zur Netzwerkforschung

1. Erstens können qualitative Verfahren wichtige Erkenntnisse liefern, wenn es darum geht, neuartige Typen von Netzwerken und Integrationsmustern zu analysieren und die relevanten Akteure sowie ihre Art der Zusammenarbeit zu erkunden.
2. Zweitens liegt die Stärke qualitativer Verfahren in der Erschließung subjektiver Wahrnehmungen, Deutungen, Relevanzen und handlungsleitender Orientierungen der Akteure eines Netzwerks. Gesamteinschätzungen und Wahrnehmungen der NetzwerkakteurInnen können auf der Basis dieser Methoden eindrucksvoll dargelegt werden.
3. Drittens kann die konkrete Netzwerkarbeit der Akteure mithilfe eines qualitativen, offenen Erhebungsdesigns rekonstruiert werden und
4. Viertens ermöglicht ein qualitatives Forschungsdesign Netzwerkveränderungen und Netzwerkdynamiken sowie den Kontext und die verursachenden Bedingungen nachvollziehbar und erklärbar zu machen.

Tabelle 14 zeigt die spezifischen Vorteile qualitativer Verfahren im Vergleich zu quantitativen Verfahren auf.

Tabelle 14: Quantitative und qualitative Verfahren der Netzwerkforschung im Vergleich

	Qualitative Verfahren	Formale Verfahren
Inhaltliche Ziele	Exploration von Netzwerken Netzwerkpraktiken Netzwerkinterpretationen Netzwerkdynamik verstehen	Netzwerkstrukturen Netzwerkdynamik
Aspekte der Datenqualität	Validität	Reliabilität Vergleichbarkeit

Hollstein 2007

Vorgehensweise der qualitativen Netzwerkforschung – Datenerhebung & Auswertung

In der Netzwerkforschung kann eine Vielzahl an unterschiedlichen qualitativen Verfahren zum Einsatz kommen. Am häufigsten werden etablierte qualitative Verfahren der Sozialforschung verwendet, wie zum Beispiel die qualitative Beobachtung, offene Interviews (narratives Interview, Problemzentriertes Interview, verstehendes Interview, Experteninterview) oder Dokumentenanalysen.

Weiters besteht die Möglichkeit der Datenerhebung durch spezifische, qualitative Netzwerkinstrumente, wie zum Beispiel die Visualisierung von Netzwerken in Netzwerkkarten oder handgezeichneten Bildern (vgl. Schulte, 1997: 24; Straus, 2002: 219), Netzwerkgraphiken oder -diagrammen (Bspw.: Methode der konzentrischen Kreise – siehe Kahn/Antonucci, 1980).

Auch bei der Datenauswertung steht der qualitativen Netzwerkforschung eine große Bandbreite an verschiedenen qualitativen Auswertungsmethoden zur Verfügung. Interaktions- Konversations- und Diskursanalysen eignen sich sehr gut, wenn es darum geht, die Netzwerkarbeit und die Netzwerkpraktiken der Akteure zu rekonstruieren. Bei der Analyse von Netzwerkwahrnehmungen kommen die verschiedensten interpretativen Auswertungsverfahren zum Einsatz. Qualitative Inhaltsanalysen oder die Grounded Theory eignen sich dafür, eher deskriptive Ergebnisse über Netzwerke zu erhalten. Möchte man die Logik des Einzelfalls nachvollziehen, werden fall-rekonstruktive Verfahren verwendet, bei denen sequenzanalytisch vorgegangen wird (Hollstein, 2007: 3365).

Verknüpfung von quantitativen und qualitativen Verfahren in der Netzwerkforschung.

Für viele Netzwerkstudien ist es sinnvoll, verschiedene Verfahren – quantitative sowie qualitative – kombiniert bei der Erforschung von Netzwerken einzusetzen. Quantitative Verfahren und Auswertungsmethoden liefern die Strukturbeschreibung, qualitative Methoden ermöglichen Aussagen über Netzwerkpraktiken, Wahrnehmungen oder die Wirkung von Kontextfaktoren. Auf diese Weise sollen „die jeweiligen Schwächen durch eine Verknüpfung der Stärken überwunden werden“ (Franke/Wald, 2006: 154).

Als ein „best-practice“-Beispiel für die Triangulation quantitativer und qualitativer Verfahren in der Netzwerkforschung kann das Forschungsdesign einer Studie von Franke und Wald (2006)¹⁰ zu Innovations- und Kooperationsnetzwerken in der Forschung angeführt werden.

In dieser Studie¹¹ wurden Forschungsgruppen aus den Disziplinen Nanotechnologie, Astrophysik und Mikroökonomie untersucht und hinsichtlich folgender Aspekte miteinander verglichen:

1. Netzwerkstrategie: Wie bauen diese Forschungsgruppen Kooperationen auf und erhalten diese?
2. Netzwerkstrukturen: Wie sind die Kooperationsstrukturen ausgestaltet?
3. Netzwerkfähigkeit: Welche Bedeutung haben Kooperationen für die Forschung?
4. Netzwerkeffekte: Welche Auswirkungen haben die Kooperationen auf den Forschungserfolg?

Die Datenerhebung erfolgte in dieser Studie durch persönliche, qualitative Leitfadeninterviews. „Für die qualitativen Elemente der Studie existiert ein Interviewleitfaden mit offenen Fragen, welche im Interview als Stimuli gesetzt wurden“ (Franke/Wald, 2006: 4399 online Text).

In den Datenerhebungsprozess wurden auch quantitative Elemente integriert, so zum Beispiel die spezifische Frage nach Kooperations- und Informationsbeziehungen sowie die standardisierte Frage nach der Intensität der Beziehungen. Zusätzlich wurden „die einzelnen Kooperations- und Informationsbeziehungen (...) vom Befragten näher kommentiert und analysiert“ (Franke/Wald, 2006: 4399 online Text).

¹⁰ Franke, Karola und Wald, Andreas (2006): Möglichkeiten der Triangulation quantitativer und qualitativer Methoden in der Netzwerkanalyse. In: Qualitative Netzwerkanalyse (Hollstein/Straus, Hg.): Konzepte, Methoden, Anwendungen. S. 153-176.

¹¹ ebd.

Die Netzwerkstruktur sowie der Forschungserfolg der Netzwerke (gemessen an der Anzahl an Publikationen) wurden durch quantitative Methoden identifiziert und ausgewertet.

Zusätzliche Informationen über die Akteursmotivation, die bewusste oder unbewusste Vorgehensweise bei der Auswahl von Kooperationspartnern sowie Kontextfaktoren für den Erfolg/Misserfolg beim Knüpfen von Netzwerken wurden mittels qualitativer Verfahren erhoben (offene Interviews) und ausgewertet (qualitative Inhaltsanalyse).

Die StudienautorInnen kommen zu folgender Conclusio: „Eine Triangulation in der Netzwerkanalyse wird unerlässlich, wenn es darum geht, nicht nur Strukturen und deren Effekte zu untersuchen, sondern auch zugrunde liegende Akteursstrategien zu identifizieren, mit denen sich Netzwerkstrukturen, -effekte und -dynamiken besser erklären und verstehen lassen“ (Franke/Wald, 2006: 4402 online Text).

Qualitative soziale Netzwerkanalyse – interessant für Wirkungsanalysen im FTI-Bereich?

In den letzten Jahren setzt die FTI-Politik verstärkt auf die Unterstützung von Innovationsnetzwerken und Clustern, um den Wissens- und Technologietransfer zu verbessern. Zusätzlich findet man den Aufbau von effizienten und flexiblen Netzwerken häufig als explizites Programmziel in FTI-relevanten Programmdokumenten wieder. Hierbei geht es um die Förderung von Innovationskapazität und Exzellenz.

Vor dem Hintergrund einer Wirkungsanalyse FTI-politischer Netzwerkiniciativen, können insbesondere folgende Fragestellungen von großer Bedeutung sein:

- Wie erfolgreich sind die neu aufgebauten/bestehenden Netzwerke?
- Welche Kontextfaktoren fördern/limitieren innovativen Output der Netzwerkakteure (Flexibilität des Netzwerks, Vertrauen, Internationalität etc.)?
- Wie kann das Innovationsverhalten der Akteure beschrieben werden? Welche Netzwerkpraktiken gibt es und welche Netzwerkkultur ist vorherrschend?
- Wie nehmen die Akteure das eigene Netzwerk vor und nach der Förderung wahr?
- Welche strukturellen Veränderungen haben sich durch die Förderung ergeben?

Um bei der Beantwortung dieser Fragen nicht bei der Beschreibung von Netzwerkinhalten und -struktur „stehen zu bleiben“, müssen zusätzlich qualitative Techniken der Netzwerkanalyse hinzugezogen werden. Das große Potential dieser Methode auch für FTI-Wirkungsanalysen wird hier augenscheinlich.

5.1.3 Aktuelle Trends in der Fokusgruppenforschung

Das Instrument der Fokusgruppe ist eine populäre und vielfach verwendete Evaluationsmethode qualitativer Sozialforschung. Fokusgruppen stellen auch ein „Standardelement“ in FTI-Evaluationen dar.

Im Gegensatz zu qualitativen Einzelinterviews handelt es sich bei der Fokusgruppe um eine moderierte und fokussierte Diskussion einer Gruppe von 6-12 Personen, die über entsprechende programmrelevante Erfahrungen und Expertise verfügen und auf der Basis ihres Wissens und ihrer

Meinung zu gemeinsamen, aber auch differenzierten Einschätzungen und Beurteilungen gelangen. Der gegenseitige Austausch und die gemeinsame Diskussion sollen in Bezug auf die eigenen Wahrnehmungen, Meinungen und Ideen befruchtend wirken. So kann es zum Beispiel zu einer Vertiefung bzw. Festigung eigener Überlegungen kommen oder aber die persönliche Meinung verändert sich im Laufe der Gruppendiskussion. Weiters werden im Diskussionsprozess der Gruppe emotionale Aspekte sichtbar, die Interaktionen der GruppenteilnehmerInnen können beobachtet und analysiert werden. All diese Aspekte können in Einzelinterviews oder aber auch durch die Aneinanderreihung mehrere Einzelinterviews nicht erreicht werden.

Gleichzeitig weist das Instrument der Fokusgruppe eine Reihe an Schwierigkeiten auf:

So hat die Auswahl der TeilnehmerInnen und deren Persönlichkeit einen wesentlichen Einfluss auf die Ergebnisse der Gruppendiskussion. MeinungsführerInnen können Gespräche dominieren, schüchterne oder zurückhaltende TeilnehmerInnen tragen keine oder nur wenige Informationen zum Thema bei; dies führt zu einer Verzerrung der Gesamtergebnisse. Weiters können Fokusgruppendiskussionen dazu führen, dass TeilnehmerInnen sensible oder persönliche Daten zurückhalten, die für den Analyseprozess bedeutsam wären. Widersprüchliche Meinungen, unklare Gruppenprozesse, das unnatürliche Setting sowie eine nur begrenzte Datenmenge beeinflussen und erschweren den Auswertungsprozess. Die international am stärksten diskutierte Schwäche von Fokusgruppen betrifft den Aspekt der „group conformity“ (Vicsek, 2010). Ergebnisse aus sozialpsychologischen Studien zu Gruppen und Gruppenverhaltensweisen zeigen, dass Menschen in Gruppen zu Konformität neigen. Das bedeutet, dass sie bei Entscheidungen in Anwesenheit Anderer den eigenen Verstand und die eigene kritische Meinung der Gruppenmehrheit unterordnen. Die Unterordnung der eigenen Meinung oder kontroverser Ideen und Standpunkte erfolgt regelmäßig auch dann, wenn die Gruppenmeinung offensichtlich falsch ist (Vicsek, 2007: 21). Dieser Druck zur Konformität muss sowohl beim Fokusgruppendesign mitbedacht als auch beim Auswertungsprozess eingeschätzt und analysiert werden.

International zeigte sich in den letzten Jahren der Trend zu Verwendung innovativer, modifizierter Fokusgruppen, um die Nachteile der traditionellen Fokusgruppenmethode bestmöglich zu vermeiden. Diese aktuellen Trends haben noch kaum Eingang in das Feld österreichischer FTI-Evaluationen gefunden und sollen daher in der Folge näher vorgestellt werden.

Aktuelle Modifikationen im Fokusgruppendesign

Two-way focus group. Bei der „two-way focus group“ beobachtet eine Gruppe von Personen – entweder verdeckt oder offen – eine andere Fokusgruppe bei der Beantwortung der Diskussionsfragen. Die Antworten der TeilnehmerInnen, ihre Interaktionen sowie der Kommunikationsprozess an sich werden von der zweiten Fokusgruppe kommentiert und diskutiert. Dieses Instrument zielt auf zwei verschiedene Vorteile ab:

- Zum einen können beide Fokusgruppenprotokolle inhaltlich ausgewertet werden. Gruppe 2 kommt auf der Basis ihrer Analyse der Inhalte von Gruppe 1 zu neuen Ideen.
- Zum anderen können Aspekte, wie z.B. Kontextfaktoren und der Aspekt der „group conformity“, die bei Gruppe 1 Einfluss auf den Gesprächsinhalt sowie Interaktionen hatten,

von Gruppe 2 analysiert werden. Im Gespräch über eine andere Gruppe fällt es Personen in der Regel leichter, noch nicht erwähnte Aspekte oder eigene Meinungen einzubringen, von denen man denkt, dass sie z.B. aufgrund „group conformity“ in Gruppe 1 nicht genannt wurden.

Mini-focus group. Bei “mini-focus groups” werden nicht 6-12 Personen, sondern maximal vier TeilnehmerInnen zu einem Diskussionsgespräch eingeladen. „Mini-focus groups“ werden insbesondere dann gerne verwendet, wenn das Thema sehr persönlich und sensibel ist. Durch die Reduktion der Gruppengröße erwartet man sich die volle Beteiligung auch von schüchternen oder zurückhaltenden TeilnehmerInnen. Zusätzlich versucht man auf diese Weise, den Einfluss der „group conformity“ zu reduzieren.

Tele-focus group. Bei “Tele-focus groups” wählen sich die TeilnehmerInnen in eine Konferenzschaltung ein und kommunizieren miteinander übers Telefon. Üblicherweise stellt ein Moderator die Leitfragen und organisiert die Diskussion; häufig werden die TeilnehmerInnen vom Moderator/ von der Moderatorin direkt gefragt, um einen Überblick darüber zu haben, wer sich bereits beteiligt hat und wer nicht. Bei „Tele-focus groups“ sind Interaktionen nur eingeschränkt analysierbar, auf die Analyse und Auswertung der Mimik und Gestik der TeilnehmerInnen muss generell verzichtet werden.

Vorteilhaft an dieser Methode ist, dass durch das Vermeiden des persönlichen Kontakts der TeilnehmerInnen größere Anonymität gewährleistet wird und es damit für die einzelnen Mitglieder einfacher ist, die eigenen Meinungen und Einstellungen vorzubringen; der Konformitätsdruck nimmt dadurch ab. Zusätzlich können „tele-focus groups“ sehr zeit- und kostensparend durchgeführt werden (Anfahrtszeiten und Reisekosten fallen weg).

E-focus group. Bei “e-focus groups”, ähnlich wie bei “tele-focus-groups”, loggen sich die TeilnehmerInnen in eine online Konferenzsoftware zu einem vorbestimmten Zeitpunkt ein (online-chatroom). Ein/eine Moderator/in leitet die Diskussion

- zum einen auf der Basis von Leitfragen, die den Kommunikationsprozess strukturieren und die wesentlichen Forschungsfragen abdecken,
- zum anderen mithilfe von visuellen Stimuli (whiteboard exercises, rating tools, smileys, etc.). Häufig werden auch webcams verwendet, um die Mimik und Gestik der TeilnehmerInnen aufzunehmen, zu interpretieren und analysieren. Vorteilhaft an dieser Methode ist der geringe Kosten- und Zeitaufwand sowie die größere Anonymität aufgrund der räumlichen Entfernung. Forschungsstudien zeigen, dass sich FokusgruppenteilnehmerInnen online stärker „öffnen“ als bei direktem „face-to-face“-Kontakt. Weiters können Ergebnisse durch die Verwendung verschiedener „Kommunikationskanäle“ erzielt werden: Über die Webcam können verbale Ergebnisse festgehalten und qualitativ ausgewertet werden, Text und visuelle Produkte (Abstimmungen, Bewertungen etc.) können quantitativ analysiert werden. Fokusgruppen können auf diese Weise sehr interaktiv und abwechslungsreich gestaltet werden, auch kleine Experimente („social experiments“) können durchgeführt werden und der Einsatz

unterschiedlicher Stimuli (wie z.B. Fotos, Videos, Texte) trägt wesentlich zur Erhaltung des Gesprächsflusses bei.

Town Hall Focus Group. „Town Hall Focus Groups“ sind große Fokusgruppen mit 20-40 TeilnehmerInnen. Zu Beginn der Fokusgruppen werden die Personen in Teams eingeteilt, ca. sechs bis acht Personen sollten ein Team bilden. Zuckerman-Parker/Shank (2008) empfehlen die Bildung von Teams auf der Basis von „Gemeinsamkeiten“ der TeilnehmerInnen (z.B. SchülerInnen-Team, LehrerInnen-Team, Eltern-Team etc.). Können sich die TeilnehmerInnen selbst nicht einem Team zuordnen, so sollten die ForscherInnen dies übernehmen, um eine zufällige Teambildung zu vermeiden – die sich in der Folge zumeist hinderlich auf den Kommunikationsprozess auswirkt. Im Verlauf des Diskussionsprozesses wird zumindest zweimal jedem Individuum die Möglichkeit gegeben, für sich als Person zu sprechen – zu Beginn der Fokusgruppe und am Ende. Dazwischen werden die Leitfragen des /der ModeratorIn gestellt und die TeilnehmerInnen diskutieren innerhalb ihres Teams (intra-group discussions) und – nachdem ein gemeinsamer Konsens gefunden wurde – zwischen den Teams (inter-group discussions). Auf diese Weise stehen den ForscherInnen sowohl Ergebnisse zum „intra team work“ als auch zum „inter team work“ zur Verfügung. Forschungsergebnisse von Zuckerman-Parker/Shank (2008) weisen daraufhin, dass sich die Teambildung positiv auf die einzelnen TeilnehmerInnen auswirkt, ihr Selbstbewusstsein stärkt und eine „stressfreie“ Diskussion ermöglicht. Dennoch sollte auch darauf geachtet werden, dass die TeilnehmerInnen am Ende einer Fokusgruppe, nachdem die Teambildung aufgelöst wurde, dazu aufgefordert werden, in einer Schlussrunde, all jene Aspekte aufzuwerfen, die bisher noch nicht erwähnt wurden. Dieser „Rollenwechsel“ gelingt nach Zuckerman-Parker/Shank (2008) sehr gut. Gestärkt aus den Gruppengesprächen fällt es den TeilnehmerInnen leichter, nun auch Aspekte zur Sprache zu bringen, die z.B. in der Gruppe „untergegangen“ sind oder in Bezug auf welche es kontroverse Teammeinungen gab.

Die Integration quantitativer Elemente in Fokusgruppen (QQM)

Der innovative Einsatz eines quantitativ-qualitativen Designs in Fokusgruppendifkussionen ermöglicht mit den Worten von Grim et al. (2006): „a collection of richer, more multifaceted data in a cost-effective fashion“ (Grim et al., 2006: 532).

Box 11: Beispiel für ein quantitativ-qualitatives Fokusgruppendedesign

Wie sieht ein quantitativ-qualitatives Fokusgruppendedesign aus?

Aktuelle Forschungsstudien (siehe Grim et al. 2006 für einen Überblick) zeigen einen Trend in folgende Richtung:

Die qualitativen Leitfadenfragen, die auch ein traditionelles Fokusgruppendedesign enthalten, bleiben bestehen; zusätzlich werden aber folgende Elemente integriert:

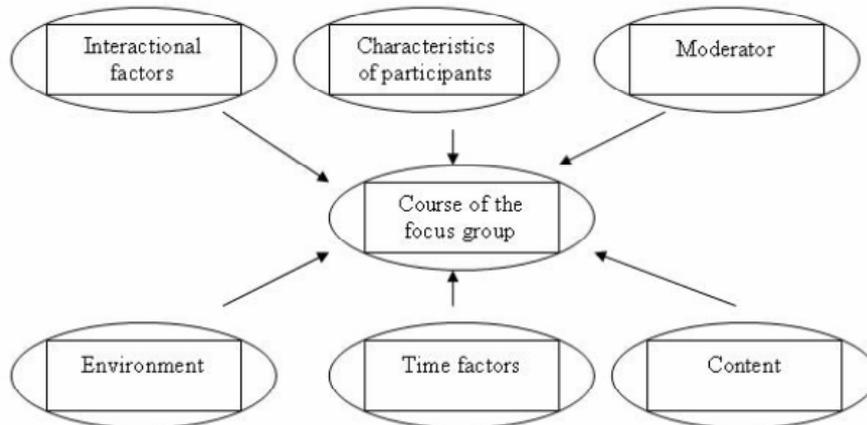
- Kurze Fragebögen mit geschlossenen Fragen, die ein aktuelles Stimmungsbild der TeilnehmerInnen zeichnen oder die Zustimmung/Ablehnung zu spezifischen Forschungsfragen erheben.
- Integration von kleinen „social experiments“: beispielsweise wird die Meinung zu einer bestimmten Frage vor der gemeinsamen Diskussion und danach erhoben und verglichen, welchen Einfluss, das gemeinsame Gespräch darauf hatte.
- Integration von Rollenspielen, bei welchen die Ergebnisse standardisiert quantitativ ausgewertet werden ; beispielsweise erhalten alle TeilnehmerInnen einen bestimmten Geldbeitrag und sie werden gefragt, wieviel dieses Geldes sie bereit wären, für einen Fahrschein zu zahlen, der bestimmte Eigenschaften ausweist? Auf diese Weise können Einstellungs- und Schwellenwerte ermittelt und analysiert werden.

Trotz der vielen Vorteile quantitativer Elemente in einer qualitativen Fokusgruppe (leichtere Datenauswertung, umfassenderes Bild der TeilnehmerInnen und ihrer Meinungen) sollte folgendem Aspekt Beachtung geschenkt werden: “quantitative activities should be kept to a proper balance so that they do not negate the greatest advantage of a focused group interview, that is its potential for generating discussions that can stimulate ideas that an individual alone may not have recollected or generated“ (Grim et al., 2006: 530).

Die Bedeutung von Kontextfaktoren bei der Auswertung von Fokusgruppenergebnissen

Sozialpsychologische Studien zeigen, dass Kontextfaktoren einen bedeutenden Einfluss auf Fokusgruppenergebnisse ausüben. Abbildung 19 zeigt die verschiedenen Faktoren (Interaktionen der TeilnehmerInnen, Charakteristika der TeilnehmerInnen, Moderator/in, Umwelt, Zeit, Inhalt) auf, die die Ergebnisse und den inhaltlichen Verlauf der Fokusgruppendediskussion lenken (können).

Abbildung 19: Potentielle Einflussfaktoren bei der Durchführung von Fokusgruppen



Vicsek 2007

International gibt es noch keinen einheitlichen Umgang in Bezug auf die Analyse von Kontextfaktoren in Fokusgruppensettings. Einige SozialwissenschaftlerInnen analysieren Fokusgruppeninterviews wie jedes andere qualitative Interview; Kontextfaktoren werden aus der inhaltlichen Analyse der Ergebnisse ausgeschlossen und Zitate werden individualisiert und nicht im Kontext der Diskussion wiedergegeben. Dies kommt daher, da diese ForscherInnen Fokusgruppenergebnisse als die Summe individueller Einstellungen und Meinungen sehen und Aspekten wie „group conformity“, „preference toward agreement“, „group norms“ und „social influence“ keine Bedeutung beimessen.

Andere ForscherInnen wiederum, darunter Lilla Vicsek, betonen die Wichtigkeit von Kontextfaktoren und erklären auf der Basis sozialpsychologischer Studien den Prozess der Meinungs- und Einstellungsformation in Gruppen. Kurz zusammengefasst können Gruppenaspekte folgende Auswirkungen haben: „Group phenomena include: less diversity of answers – because of preference toward agreement, some people cease to talk about their opinion when they find out that other represent a different view and group situations might influence certain individual experiences“ (Vicsek, 2010: 134).

Wie kann damit umgegangen werden? Vicsek (2010) schlägt folgendes Auswertungsschema vor:

1. Zitate immer im Kontext der Diskussion wiedergeben, wenn möglich auch Kommentare in Klammer setzen (z.B. alle lachen, ein „Raunen“ geht durch den Raum etc.)
2. Bei der Auswertung und Analyse der Ergebnisse folgendermaßen vorgehen:
 - a Kontextfaktoren nennen und deren möglichen Einfluss auf die Ergebnisse beschreiben, e.g.
 - the participants of the groups were each other’s colleagues,
 - some of the groups were homogenous with respect to gender, while others contained both men and women,

- the head of the work organisation did the recruitment, he asked the research participants to take part in the focus groups,
 - the scene of the discussions was a room at the workplace,
 - the conversations were video-recorded,
 - the moderator was a woman,
 - the atmosphere – except for one group – was relaxed,
 - the topic did not cause great conflicts among group members
- (Vicsek 2010: 134)

b Inhaltliche Ergebnisse auswerten

c Ergebnisse der Kontextanalyse und der thematischen Analyse zusammenführen und für jedes Hauptergebnis festlegen, welche Kontextfaktoren in welcher Art und Weise Einfluss ausgeübt haben könnten (Ergebnis wird verstärkt oder aber abgeschwächt), Schlussfolgerungen formulieren

3. So genannte „screening questionnaires“ zum Abschluss der Fokusgruppe austeilen mit 1-2 offenen Fragen, die die TeilnehmerInnen auffordern, die eigene Meinung zur Hauptforschungsfragestellung zu formulieren sowie Nachbesprechungen (durch Angabe eines Telefonkontakts oder einer Email-Adresse) anbieten, damit die TeilnehmerInnen sich beim Forscher/der Forscherin melden können, sofern ihnen noch weitere Gedanken kommen/Aspekte einfallen.

5.2. QUANTITATIVE METHODEN

Es gibt zahlreiche Methoden, die bei der Evaluation von innovationspolitischen Maßnahmen angewandt werden und die als „quantitative“ Methoden bezeichnet werden können. Typische Beispiele sind die quantitative Netzwerkanalyse oder Input-Output Analysen. Allerdings gibt es nur einen quantitativen, methodischen Ansatz, der die Kernfrage jeder Evaluation in den Mittelpunkt des Erkenntnisinteresses stellt. Dies ist eine „Was-wäre-wenn“ Frage und lautet: Wie hätte sich das österreichische Innovationssystem (oder irgendein anderer Ausschnitt der sozialen Realität) entwickelt, wenn nicht eine bestimmte politische Intervention stattgefunden hätte? Erst die Antwort auf diese Frage ermöglicht es, kausale Effekte der Politik abzuschätzen.

Die Disziplin, die sich dieser Frage widmet, ist die Mikroökonometrie, die im Englischen als „econometrics of program evaluation“ (Wooldridge/ Imbens 2009) oder „treatment evaluation“ (Cameron/ Trivedi 2005) bezeichnet wird. Zentrales Kennzeichen dieser Literatur ist das Streben nach einer methodisch abgesicherten Kausalanalyse..

Generell ist anzumerken, dass mikroökonometrische Modelle einen hohen Grad an Komplexität aufweisen und solide ökonometrische und statistische Kenntnisse voraussetzen. Daher werden in diesem Kapitel nicht die detaillierte Herleitungen von Modellen diskutiert oder die vielfachen, durchwegs subtilen Argumentationslinien in ihren Einzelheiten nachgezeichnet. Es sei hier nur erwähnt, dass oftmals aufgrund mathematischer Darstellungsweisen ein Eindruck von Objektivität

und „Wahrheit“ vermittelt wird, der über die zum Teil intensiven Debatten innerhalb der scientific community hinwegtäuscht. Tatsächlich sind eben viele methodischen Fragen noch keineswegs abschließend gelöst.

Eine Fokussierung der Darstellung auf mikroökonomische Methoden kann wie folgt begründet werden: Erstens ist aus pragmatischen Gründen eine Einschränkung der Stoffauswahl notwendig. Zweitens versucht die Mikroökonomie explizit kausale Fragestellungen zu beantworten und erfüllt damit ein wichtiges Bedürfnis von Evaluation. Drittens ist festzustellen, dass die in der mikroökonomischen Literatur eingeführten Kategorien einen hilfreichen Denkraum auch für andere Evaluierungsmethoden bieten. Die exakte Formulierung der mit Kausalanalysen verbundenen Problemstellungen schafft eine Sensibilität zur Einschätzung der Rolle kritischer Annahmen in Evaluierungen und erhöht das Kritikpotenzial auf Seiten der Konsumenten sozialwissenschaftlicher Arbeiten. Viertens wurde in den vorangehenden Kapiteln auch eine Lücke bei der Anwendung ökonomischer Arbeiten zur Evaluation innovationspolitischer Maßnahmen in Österreich argumentiert. Daher erscheint es sinnvoll, diese Modelle hier in den Vordergrund zu stellen.

Das Ziel dieser Darstellung ist zweifach. Neben der Präsentation eines allgemeinen Analyserahmens für die Formulierung kausaler Fragestellungen werden die gebräuchlichsten mikroökonomischen Modelle vorgestellt. Die Darstellung beschränkt sich weitgehend auf die Diskussion der kritischen Annahmen, auf denen diese Modelle aufbauen.

Dieses Unterkapitel ist wie folgt gegliedert: Zunächst erfolgt in Abschnitt 5.2.1 eine kurze Darstellung der Entwicklung der Mikroökonomie und ihrer zunehmenden Bedeutung in den letzten Jahrzehnten. In Unterkapitel 5.2.2 wird der allgemeine, konzeptionelle Denkraum zur ökonomischen Evaluation präsentiert. Die grundsätzlichen Strategien zur Lösung des Evaluationsproblems werden in Abschnitt 5.2.3 diskutiert. Dabei wird auch die häufig angewandte Methode der „direkten Frage“ nach kausalen Effekten erörtert und ihre Problematik vor dem Hintergrund des hier vorgestellten analytischen Zugangs betont. In Abschnitt 5.2.4 werden einige der am häufigsten verwendeten ökonomischen Methoden dargestellt und mit Anwendungsbeispielen illustriert. Grenzen ökonomischer Modellierung werden in Abschnitt 5.2.5 thematisiert.

5.2.1 Entwicklung der mikroökonomischen Programmevaluierung

Mikroökonomie als Methode bezeichnet die Verbindung von mikroökonomischer Theorie und Ökonometrie bzw. Statistik. Besonderes Kennzeichen ist die Verwendung von Mikrodaten, d.h. Daten auf Ebene der einzelnen ökonomischen Akteure wie Haushalte (Individuen) und Unternehmen. Nach Cameron und Trivedi (2005:3) versteht man unter Mikroökonomie “the analysis of individual-level data on the economic behaviour of individuals or firms.” Während lange Zeit vorwiegend Makrodaten analysiert wurden, ermöglichte die zunehmende Verfügbarkeit von Mikrodaten sowie die weitere Verbesserung der Rechenkapazitäten ab den 1960er Jahren die Entwicklung der Mikroökonomie. Man beachte hier die „Datengetriebenheit“ der methodischen Entwicklung. Die Frage der Daten ist demnach kein Nebenschauplatz sondern zentraler und

immanenter Bestandteil empirischer und methodischer Forschung. Eine wichtige Ursache für die verzögerte Anwendung mikroökonomischer Methoden in der Innovationsökonomik ist die Tatsache, dass Mikrodaten zunächst nur für den Haushaltssektor zu Verfügung standen. Erst später standen auch Mikrodaten für Unternehmen zur Verfügung, die bei der Evaluation innovationspolitischer Maßnahmen freilich von entscheidender Bedeutung sind. Einen vorläufigen Höhepunkt der Mikroökonomie als Wissenschaft stellte der Nobelpreis an die Mikroökonomiker Heckman und McFadden im Jahr 2000 dar; James Heckman wurde auch aufgrund seiner Arbeiten zur Programmevaluation ausgezeichnet.¹²

Wie bereits erwähnt ist die Literatur zur Programmevaluation ein Teilbereich einer umfassenderen mikroökonomischen Disziplin. Nicht zuletzt durch die säkular steigenden Staatsquoten in vielen Staaten sowie die zunehmende Schulden wurden Methoden zur Identifikation effektiver politischer Intervention wichtiger. Aus theoretischer Perspektive sind nach Verbeek (2008) zwei Faktoren zu nennen, die eine Beschäftigung mit der Frage von Maßnahmeneffekten interessant machen: Erstens sind die Maßnahmeneffekte für die verschiedenen Maßnahmenteilnehmer unterschiedlich. Zweitens ist die Selektion in ein Programm kein Ergebnis des Zufalls, sondern von bewussten Entscheidungen verschiedener ökonomischer Akteure. All dies trug dazu bei, dass Greene (2011:930) mittlerweile von einer „huge and rapidly growing literature“ sprechen kann. Auch Schmidt (2007: 1,6) konstatiert eine Bedeutungszunahme und vor allem auch eine Erkenntniszunahme in diesem Forschungsfeld: „Arguably, one of the most important developments in the field of applied economics during the last decades has been the emergence of systemic policy evaluation, with its distinct focus on the establishment of causality. (...) It is now up to the demand side to make full use of this potential for the design of better economic policy.“ Umfassende Literaturüberblicke, etwa von Wooldridge und Imbens (2009) oder von Blundell und Dias (2008) dokumentieren eine Konsolidierung dieses methodischen Forschungszweigs.

Wichtige Weiterentwicklungen seit den Anfängen der Programmevaluierung¹³ betreffen vor allem zwei Bereiche: Erstens wurden anfangs hochparametrische Modelle zunehmend durch semi- und nicht-parametrische Modelle ersetzt.

Diese erfordern weniger oder keine Annahmen über Verteilungsformen (z.B. Annahme der Normalverteilung), funktionale Spezifikationen (z.B. lineare Gleichungsmodelle) oder Ausschlussrestriktionen (z.B. Annahmen über die Nichtberücksichtigung von Variablen). Damit werden die Ergebnisse tendenziell robuster. Zweitens setzte sich die Erkenntnis durch, dass das Studiendesign von entscheidender Bedeutung für die Validität kausaler Schlussfolgerungen ist. Diese Entwicklung wurde vor allem durch Angrist und Pischke (2010, 2009) geprägt und mit dem Begriff „Credibility Revolution“ in die Literatur eingeführt. Um deren Argumente zu verstehen ist es notwendig, die Situation der ökonomischen Forschung in den 1980er Jahren zu betrachten. Den damaligen, unbefriedigenden Stand der Debatte fasste Leamer (1983:37) mit folgendem Statement zusammen: „Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analysis serious.“ Wenig später veröffentlichte Lalonde

¹² http://www.nobelprize.org/nobel_prizes/economics/laureates/2000/# [25.11.2011].

¹³ Im Folgenden wird unter „Programmevaluierung“ stets eine mikroökonomische Programmevaluierung verstanden.

(1986) einen Aufsatz der weitreichende Konsequenzen haben sollte. Darin kommt er zum Schluss, dass ökonometrische Evaluierungen von Programmen der aktiven Arbeitsmarktpolitik auf der Basis von historischen Daten – Daten, die nicht durch Experimente generiert wurden – keine Replikation von Ergebnissen sozialer Experimente ermöglichen. Nachdem die Glaubwürdigkeit von Experimenten generell hoch eingeschätzt wird, nahmen die Zweifel an ökonometrischen Evaluierungen weiter zu. Schließlich setzte sich die Meinung durch, dass auch ökonometrische Studien mit historischen Daten versuchen sollten, in der einen oder anderen Form experimentelle Forschungsdesigns nachzuahmen. Es lässt sich also feststellen, dass insgesamt das soziale Experiment deutlich an Bedeutung zur Programmevaluation gewonnen hat und gleichzeitig auch nicht-experimentelle Methoden sich an dem Ideal eines randomisierten Experiments orientieren. Typisches Beispiel hierfür ist das noch näher zu erörternde Matching-Verfahren.

Der Einsatz mikroökonometrischer Methoden zur Evaluierung von F&E-Programmen ist noch nicht so weit verbreitet wie in anderen Politikfeldern (z.B. Arbeitsmarkt- oder Bildungspolitik). Trotzdem wurde bereits in Kapitel 0 gezeigt, dass die Anzahl der Anwendungen keineswegs vernachlässigbar sondern sogar deutlich im Zunehmen ist. Auch hier können sicherlich Fragen der Datenverfügbarkeit als Erklärung dienen. Im Wesentlichen handelt es sich dabei meist um die Anwendung von Methoden, die ursprünglich für arbeitsmarktökonomische Fragestellungen entwickelt wurden, auf innovationsökonomische Anwendungsfälle. Wichtige, genuin innovationsökonomische Fragestellungen wurden ab den 1970er Jahren im Rahmen von Untersuchungen zur Spillovers untersucht. Nachdem die Existenz von positiven externen Effekten (Knowledge Spillovers) eine wesentliche Begründung für innovationspolitische Intervention darstellt, kommt diesen Untersuchungen eine wichtige Bedeutung zu. Klette et al. (2000) zeigen hierzu einen nach wie vor aktuellen Überblick. Erstaunlicherweise ist die Anzahl von solchen Untersuchungen zu Spillovers aber sehr gering. Die zweite wesentliche Innovation war die Übernahme der Erkenntnis aus der Evaluation von Arbeitsmarktmaßnahmen, dass die Selektion von geförderten Unternehmen im Rahmen eines F&E-Programms gravierende Auswirkungen auf die Evaluierung dieses Programms hat. Dieser unter dem Stichwort Sample-Selection Bias oder Endogenitätsproblem debattierte Sachverhalt wird weiter unten noch näher erörtert. Vorläufig sei jedoch festgehalten, dass ein Großteil der Forschung in der einen oder anderen Form eine Antwort auf dieses Problem darstellt. Die erste bedeutende Arbeit, die dieses Problem in einem F&E-Kontext erörterte, ist Lichtenberg (1987).

Die zunehmende Anwendung mikroökonometrischer Programmevaluierung wird auch bei der 2003 vom deutschen Bundesministerium für Bildung und Forschung beauftragten Studie „Einsatzmöglichkeiten neuer quantitativer Ansätze zur Wirkungsanalyse öffentlicher Forschungsförderung. Eine Grundlagenstudie“ deutlich. Diese von Czarnitzki et al. (2003) verfasste Studie analysiert die Anwendungsmöglichkeiten und Probleme ökonometrischer Evaluierungen anhand eines umfassenden Mikrodatensatzes. Eine aktuelle systematische Zusammenschau der bisherigen mikroökonometrischen FTI-Evaluierungen bietet Cerulli (2010) auf einem methodisch anspruchsvollen Niveau. Zusammenfassend zeigt sich im Bereich Innovationsökonomie bzw.

Innovationspolitik eine Integration ökonomischer Evaluierungsmethoden in den Mainstream der Forschung und Evaluierung innerhalb der letzten beiden Jahrzehnte.

5.2.2 Rubin Causal Model und das Evaluationsproblem

Die Formulierung kausaler Fragestellungen erfolgt in der Mikroökonomie in einer modellhaften, abstrakten Form. Dieses „Rubin Causal Model“ wurde nach dem Statistiker Donald Rubin benannt, der in den 1970er Jahren die intellektuellen Grundlagen erarbeitete.

Zur besseren Veranschaulichung gehen wir von einem idealtypischen Programm zur Förderung unternehmensbezogener F&E aus, etwa im Sinne der FFG-Einzelprojektförderung im Rahmen der Basisprogramme. Wir nehmen weiters an, dass die Innovationsförderung das Ziel verfolgt, den Patent-Output der Unternehmen zu erhöhen, da es sich hierbei um einen wichtigen Indikator technologischer Leistungsfähigkeit und um einen quantitativ und objektiv erfassbaren Parameter handelt. In technischen Termini ausgedrückt handelt es dabei um die Erfolgsgröße. Freilich ist das eine bewusst einfach gewählte, eindimensionale Erfolgsgröße.

Mikroökonomische F&E-Evaluierungen analysieren in der Regel die Programmwirkungen auf mehrere Erfolgsgrößen, z.B. werden neben Patenten auch der Umsatz mit innovativen Produkten oder die Entwicklung der F&E-MitarbeiterInnenzahlen berücksichtigt. Qualitative Erfolgsgrößen können etwa durch mehrere quantitative Erfolgsgrößen operationalisiert werden. Im Allgemeinen werden die Erfolgsgrößen durch die in den Programmdokumenten vorgegebenen Ziele determiniert. Es wäre aber auch wichtig, dass bei der Formulierung der Ziele die Verfügbarkeit von möglichen Ergebnisgrößen mitgedacht wird, jedenfalls wenn eine ökonomische Evaluierung angedacht wird.

Das Evaluationsproblem

Ausgangspunkt des Rubin Causal Model ist die Vorstellung, dass ein Unternehmen am F&E-Förderprogramm potenziell teilnehmen oder nicht teilnehmen kann, daher der Begriff „potenzielle Ergebnisse“. Das Ergebnis eines Unternehmens i nach der Teilnahme am innovationspolitischen Programm wird mit $Y_i(1)$ bezeichnet, während $Y_i(0)$ das Ergebnis bei Nichtteilnahme bezeichnet. Der einzige Unterschied zwischen diesen beiden Zuständen ist demnach die Maßnahmenteilnahme. Der kausale Effekt der Förderung, Δ_i , ergibt sich also gemäß folgender Gleichung:

$$\Delta_i = Y_i(1) - Y_i(0). \quad (1)$$

Das offensichtliche Problem ist jedoch, dass nur eines der beiden potenziellen Ergebnisse realisiert werden kann. Ein Unternehmen erhält eine Förderung oder erhält keine Förderung, beides gleichzeitig ist nicht möglich. Deswegen ist Gleichung (1) nicht bestimmt, oder in technischen Termini nicht identifiziert. Ein Parameter gilt dann als identifiziert, wenn er bei stetig anwachsender Stichprobengröße mit einer immer größeren Genauigkeit geschätzt werden kann (Bauer et al. 2009). Das bedeutet, dass selbst bei einer unendlich großen Stichprobe Δ_i nicht geschätzt werden kann. Auch ein Experiment kann die notwendige Information nicht generieren.

Dieses Problem wird als das „fundamental problem of causal inference“ oder einfach als „Evaluationsproblem“ bezeichnet. Es ist letztlich das Ergebnis fehlender und nicht ermittelbarer Information.

Es muss bereits an dieser Stelle betont werden, dass die Situation ohne Förderung nicht einfach einer Situation ohne F&E-Projekt entspricht. Tatsächlich wird dies eher die Ausnahme darstellen. Vielmehr ist etwa damit zu rechnen, dass Unternehmen die keine Förderung erhalten zahlreiche kreative Wege ausprobieren werden um trotzdem Innovationsaktivitäten zu setzen.

Das realisierte Ergebnis der beiden potenziellen Ergebnisse ergibt eine Funktion der Maßnahmenteilnahme. Im Folgenden bezeichnet D_i eine Indikatorvariable, die den Wert 1 annimmt wenn Unternehmen i an der Maßnahme teilnimmt, und den Wert 0, wenn das Unternehmen nicht gefördert wird. Dann ergibt sich das tatsächlich realisierte und damit beobachtbare Ergebnis als

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) \quad (2)$$

Damit ist eine Unterscheidung zwischen potenziellen und realisierten Ergebnissen getroffen. Je nachdem ob $D_i=1$ oder $D_i=0$ wird eines der beiden potenziellen Ergebnisse zum realisierten Ergebnis. Das jeweils andere Ergebnis bildet die sogenannte kontrafaktische Situation, die sich vor allem dadurch auszeichnet, dass sie nicht beobachtet werden kann (Tabelle 15).

Tabelle 15: Beobachtbare und unbeobachtbare Situationen

	$D_i=1$	$D_i=0$
$Y_i(1)$	Beobachtbar	Nicht beobachtbar (Kontrafaktische Situation)
$Y_i(0)$	Nicht beobachtbar (Kontrafaktische Situation)	Beobachtbar

Eigene Darstellung nach Bauer et al. 2009

Um ein illustratives Beispiel zu geben nehmen wir an ein Unternehmen erhält eine Innovationsförderung und hat zwei Jahre nach Programmende zwei Patente angemeldet. Um den Programmeffekt zu bestimmen, müssten wir wissen, wie viele Patente das Unternehmen ohne Förderung angemeldet hätte. Diese kontrafaktische Situation ist jedoch nicht beobachtbar und damit kann diese Fragestellung grundsätzlich nicht beantwortet werden. Offensichtlich ist eine Lösung des Evaluationsproblems, d.h. eine numerische Lösung von Gleichung (1) auf Ebene einzelner Unternehmen nicht möglich. Stattdessen versucht man die Lösung über die Ermittlung von unterschiedlichen Durchschnitten zu ermitteln.

Dies führt zu der Frage nach dem zu verwendenden Evaluationsparameter.

Evaluationsparameter

Die Aufgabe von Evaluationsparametern ist es, die kausalen Effekte der Maßnahme in geeigneter Art und Weise darzustellen bzw. zusammenzufassen. Dabei stehen mehrere unterschiedliche

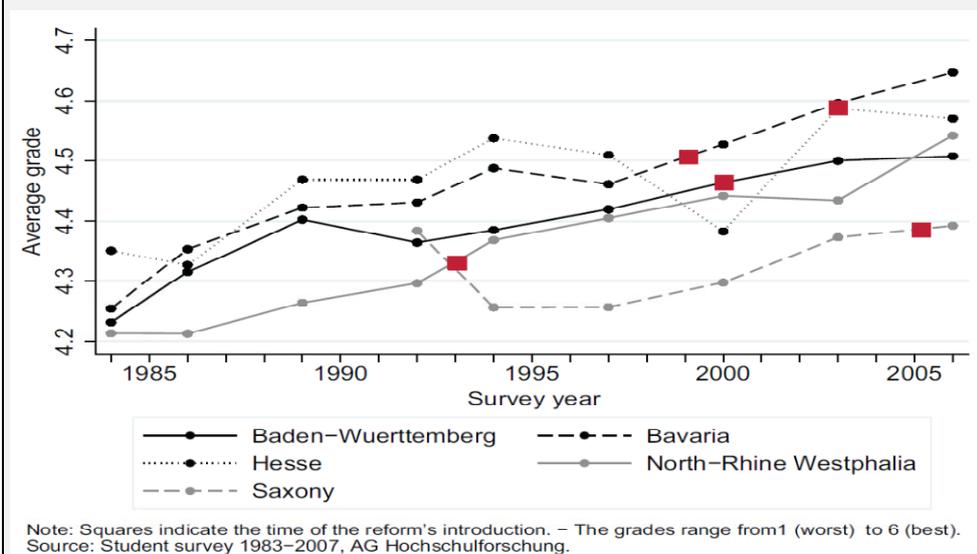
Evaluationsparameter zur Auswahl. Bedeutend ist die Entscheidung für einen Parameter vor allem vor dem Hintergrund der Frage, in wie weit innovationspolitische Maßnahmen auf unterschiedliche Typen von Unternehmen unterschiedliche Effekte haben, etwa z.B. ob der Einfluss der Förderung auf Unternehmen mit geringen Patentaktivitäten höher ist als auf Unternehmen, die viele Patente anmelden. Ist dies der Fall so spricht man von heterogenen Maßnahmeneffekten. Ist der kausale Programmeffekt für alle Unternehmen gleich, bezeichnet man die Maßnahmeneffekte als homogen. Trifft Letzteres zu ist die Wahl des Evaluationsparameters belanglos, weil dann alle Evaluationsparameter das gleiche Ergebnis liefern. Generell lassen sich Evaluationsparameter danach unterscheiden, ob sie in der Lage sind die Heterogenität der Programmwirkung zu erfassen oder ob sie lediglich einen Durchschnittswert für alle Teilnehmer liefern. Am häufigsten werden in Evaluierungen Durchschnittsparameter verwendet. Dies liegt nicht zuletzt daran, dass Evaluationsparameter, welche die Verteilung der Effekte erfassen, komplexer sind. Tatsächlich markiert diese Fragestellung eine der Forschungsfronten im Bereich der Programmevaluierung (Wooldridge/ Imbens 2009). Freilich wird mit der Verwendung von Parametern zur Erfassung durchschnittlicher Maßnahmeneffekte eine gewisse Ungenauigkeit in der Aussage hingenommen. Je heterogener die Maßnahmeneffekte, desto fragwürdiger wird dieser Zugang. Ein Beispiel für die Analyse von Effekten einer Maßnahme auf unterschiedliche Momente der Verteilung der Ergebnisgrößen gibt Box 12 anhand des Evaluationsparameters „Quantile Treatment Effects“.

Box 12: Erfassung heterogener Maßnahmeneffekte. Fallbeispiel Universitätspolitik

Bauer und Grawe (2011) evaluieren in ihrer Studie die Folgen der Einführung eines indikatorbasierten Finanzierungssystems in einigen deutschen Bundesstaaten. Dieses soll den Universitäten Anreize bieten, um zu effizienten und ergebnisorientierten Organisationen zu werden. Die Kräfte des Marktes sollten auch im Hochschulsektor für effiziente Mittelverwendung sorgen. Entsprechend wird ein Teil der Mittelzuweisung anhand von Leistungsindikatoren vergeben. Allerdings ist in der innovationsökonomischen Literatur bekannt, dass eine solche Orientierung an quantitativen Indikatoren eine Reihe von negativen, nicht-intendierten Effekten haben kann. Nachdem auch die Abschlusszahlen von StudentInnen ein Leistungsindikator sind, ist zu befürchten, dass es zur Noteninflation kommt, d.h. die Universitäten maximieren ihre AbsolventInnenzahlen und damit ihr Einkommen, indem sie möglichst vielen StudentInnen den Abschluss ermöglichen. Es ist klar, dass solch ein Ergebnis aus volkswirtschaftlicher Perspektive unerwünscht ist und demnach ist es wichtig festzustellen, ob die Universitätsreformen in einigen Bundesländern solche adversen Anreizeffekte verursachten.

Die Datenbasis der Studie bilden die regelmäßig stattfindenden StudentInnenbefragungen in Deutschland, wobei hier eine Zeitreihe von 1983-2007 zur Verfügung stand. Jede Befragungswelle beinhaltet etwa 7.000-10.000 StudentInnen. Die Hochschulen des Bundeslands Nordrheinwestfalen werden als Maßnahmenteilnehmer verwendet, während jene von Bayern und Baden-Württemberg als Kontrollgruppe dienen. Abbildung 20 zeigt die Entwicklung der Durchschnittsnoten für ausgewählte deutsche Bundesländer. Dabei fällt auf, dass in allen Bundesländern eine Verbesserung des durchschnittlichen Notenniveaus über die Zeit stattgefunden hat. Offensichtlich lässt sich der Effekt der Reform aber nicht einfach aus einer visuellen Introspektion der Zeitreihen ableiten. Die Autoren nutzen bei ihrer Identifikationsstrategie die Tatsache, dass Nordrheinwestfalen die Universitätsreform mehrere Jahre vor den anderen Bundesländern einführte.

Abbildung 20: Entwicklung der durchschnittlichen Noten von StudentInnen



Bauer/Grave 2011

Durch diese zeitliche Struktur sowie die Verfügbarkeit von Paneldaten ist die Anwendung eines Differenz-von-Differenzen Ansatzes möglich (siehe unten). Dieser ermöglicht sowohl die Abschätzung durchschnittlicher kausaler Effekte als auch von Verteilungseffekten der Reform auf die Notenverteilung. Kausale Verteilungseffekte werden durch den Evaluationsparameter „Quantile Treatment Effects“ (QTE) erreicht, der den kausalen Reformeffekt auf die unterschiedlichen Quantile der Notenverteilung angibt. Damit kann die Hypothese untersucht werden, ob alle StudentInnen bessere Noten aufgrund der Reform erhalten oder ob etwa nur die Barrieren zum positiven Abschluss einer Prüfung reduziert wurden. Im ersten Fall kommt es zu einem kausalen Programmeffekt auf alle Quantile mit Ausnahme des Quantils mit jenen StudentInnen, die bereits die bestmögliche Note haben. Im letzteren Fall kommt es zu einer Kompression der Notenverteilung und die Reform verursacht eine Noteninflation lediglich in den unteren Quantilen. Konkret analysieren Bauer und Grave die Effekte für das 10., 25., 50., 75. und 90. Quantil.

Die Ergebnisse der Evaluierung zeigen, dass die Reform keine Noteninflation verursacht und auch keine spezifischen Verteilungseffekte auf das Notenspektrum hatte. *Bauer/Grave 2011*

Die weitere Darstellung mikroökonomischer Methoden folgt der gängigen Darstellung in der einschlägigen Literatur und konzentriert sich auf genau einen Evaluationsparameter, nämlich den sogenannten durchschnittlichen Maßnahmeneffekt auf die Maßnahmenteilnehmer (mean effect of treatment on the treated, MTT). Er liefert die Antwort auf die Frage: „Wie groß ist der durchschnittliche kausale Maßnahmeneffekt eines geförderten Unternehmens im Vergleich zur Situation ohne Förderung?“ Der MTT ist der gebräuchlichste Evaluationsparameter. Um den MTT näher zu beschreiben, muss zunächst die Notation sogenannter bedingter Erwartungswerte eingeführt werden. Darunter versteht man den Erwartungswert einer Zufallsvariable, wenn eine andere Zufallsvariable einen bestimmten Wert annimmt. Der Erwartungswert bezieht sich dabei auf die gesamte Population und wird durch das arithmetische Mittel aus Stichproben geschätzt. Die formale Notation für einen bedingten Erwartungswert ist $E(Y|X)$. Darunter versteht man also den Erwartungswert von Y bedingt auf X. Der MTT ergibt sich demnach aus folgender Gleichung:

$$MTT = E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 1). \quad (3)$$

Dabei bezeichnet der erste Term auf der rechten Seite von Gleichung (3) die durchschnittliche Ergebnisgröße, d.h. in unserem Fall die durchschnittliche Anzahl angemeldeter Patente bei Teilnahme des Unternehmens i an der Fördermaßnahme, während der zweite Term den bedingten Erwartungswert der Erfolgsgröße bei Nichtteilnahme zeigt. Offensichtlich ist das Ergebnis bei Nichtteilnahme ($Y_i(0)$) bedingt auf die Teilnahme ($D_i=1$) kein in der Realität beobachtbarer Zustand. Damit stellt sich erneut das Problem, dass auch Gleichung (3) und damit der kausale Maßnahmeneffekt nicht identifiziert ist, weil der zweite Term auf der rechten Seite eine kontrafaktische Situation bezeichnet. Weil aber jetzt nicht mehr auf der Ebene eines einzelnen Unternehmens argumentiert wird, kann die kontrafaktische Situation durch eine alternative,

beobachtbare Situation ersetzt werden. Zur Lösung dieses Problems gibt es verschiedene, sogenannte Identifikationsstrategien, die im nächsten Unterkapitel dargestellt werden. Zuvor soll jedoch noch eine kritische Annahme erörtert werden, die in diesem Abschnitt bislang implizit geblieben ist, aber konstitutiv für mikroökonomische Analysen ist. Es handelt sich hierbei um die sogenannte Stable-Unit-Treatment-Value-Asumption.

Stable-Unit-Treatment-Value-Asumption (STUVA)

Darunter versteht man die Annahme, dass die Teilnahme eines Unternehmens an einem innovationspolitischen Förderprogramm keine Effekte auf die Nichtteilnehmer oder andere Teilnehmer hat. Entscheidend für die Programmeffekte ist ausschließlich der jeweils unternehmensspezifischen Förderstatus. Aus diesem Grund spricht man auch von der „Lack-of-Interaction-Assumption“. Diese Annahme ist etwa in der medizinischen Forschung oftmals erfüllt. So ist z.B. der Erfolg der Behandlung eines/r KrebspatientIn mit einem neuen Medikament unabhängig davon, ob andere dieses Medikament erhalten oder nicht. Bei der Behandlung von epidemischen Krankheiten wäre aber z.B. die STUVA-Annahme verletzt.

Bezogen auf innovationspolitische Evaluation ist diese Annahme durchaus kritisch, weil die Legitimation innovationspolitischer Intervention ja gerade in der Existenz von Spillovers besteht (Fritsch et al. 2007).

Generell würde die Existenz von Spillovern zu einer Unterschätzung der Maßnahmeneffekte führen, da dann auch die Nichtteilnehmer als Vergleichsgruppe über „Umwege“ von der Förderung profitieren. Dieses Problem ist Gegenstand methodischer Forschung (Czarnitzki et al. 2003) und nicht zuletzt deswegen von zunehmender Relevanz, weil F&E-Spillovers aufgrund technologischer Entwicklungen in den letzten Jahrzehnten an Bedeutung gewonnen haben (Badinger 2011). Aber auch Verdrängungseffekte im Sinne negativer Externalitäten könnten eine wichtige Rolle spielen. So ist es etwa durchaus vorstellbar, dass geförderte Unternehmen nicht geförderte Unternehmen vom Markt verdrängen. Diese negativen Effekte bleiben bei einer mikroökonomischen Analyse grundsätzlich unberücksichtigt. Ein Beispiel aus dem Bericht des deutschen Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung illustriert die Problematik anhand von industriepolitischen Maßnahmen in der Wirtschaftskrise.

Box 13: Rückwirkung industriepolitischer Maßnahmen auf andere Unternehmen – ein stilisiertes Beispiel

Während sich beispielsweise in der Medizin häufig ein positiver Nebeneffekt einstellt – wer aufgrund einer Impfung nicht krank wird, steckt daher andere nicht an –, dürften bei industriepolitischen Maßnahmen eher negative Begleiterscheinungen zu befürchten sein: Angenommen sei zum Beispiel, dass eine Stadt 100 Dienstfahrzeuge zu erwerben plant, davon 50 der Marke A und 50 der Marke B. Aufgrund einer Krise bei Hersteller A entscheidet sich die Stadt, nunmehr 100 Fahrzeuge der Marke A zu erwerben – die Beschäftigung bei Hersteller A wird stabilisiert. Dies ist aber wiederum nur ein Bruttoeffekt, denn es wäre fahrlässig, den Umstand zu ignorieren, dass Hersteller B einen Rückgang seines Absatzes hinnehmen und gegebenenfalls ArbeitnehmerInnen entlassen muss. Der Nettoeffekt ist sogar negativ wenn man berücksichtigt, dass sich die Stadtverwaltung wohl nicht ohne Grund für die ursprünglich hälftige Aufteilung der Fahrzeuge entschieden *hatte*.

Sachverständigenrat Jahresgutachten (2009/10: 230).

Letztlich manifestiert sich in dieser Annahme die Tatsache, dass mikroökonomische Modelle auf der Vorstellung eines partiellen Gleichgewichts beruhen und kein allgemeines Gleichgewichtsmodell gleichsam automatisch „im Hintergrund“ mitgerechnet wird. Damit werden aber auch keine Interdependenzen zwischen den Märkten und Akteuren berücksichtigt und allgemeine Gleichgewichtseffekte unberücksichtigt gelassen. Diese können etwa auftreten wenn eine Maßnahme einen substanziellen Teil des Unternehmenssektors eines Staates oder einer Region betrifft. Es könnte z.B. folgender, durchaus realistische Fall eintreten: Eine innovationspolitische Maßnahme wird zunächst in einer Pilotregion eingeführt. Nachdem die Evaluation positive Effekte auf den Absatz der Unternehmen gezeigt hat, wird die Maßnahme nunmehr auf die gesamte Volkswirtschaft ausgeweitet. Weil aber der relevante Markt im Aggregat für die geförderten Unternehmen weitgehend gesättigt ist, hat die Maßnahme bei breiter Implementierung wesentlich geringere Effekte als dies im Pilotstadium der Fall war. Hier konnten die geförderten Unternehmen ihren Absatz auf Kosten der nicht-geförderten Unternehmen ausdehnen.

Damit wird die wichtige Funktion von allgemeinen Gleichgewichtsmodellen deutlich. Diese sind – jedenfalls theoretisch – in der Lage, alle Interdependenzen zwischen den Märkten zu berücksichtigen. Ein Beispiel für eine F&E-bezogene Anwendung ist das australische Monash-Modell, ein dynamisches allgemeines Gleichgewichtsmodell, welches zur Evaluation der Förderungen des „Australia Research Council“ verwendet wird (OECD 2008).

5.2.3 Fundamentale Identifikationsstrategien und Sample Selection Bias

Die Diskussion des Evaluationsproblems hat deutlich gemacht, dass die Beantwortung der Frage nach kausalen Programmeffekten nicht ohne weiteres möglich ist, weil sie die Formulierung einer kontrafaktischen Situation notwendig macht, die nicht beobachtet werden kann. Damit ist zunächst einmal klar, dass eine exakte Bestimmung oder gar Messung von Programmeffekten unmöglich ist. Das ist bereits eine sehr wichtige Erkenntnis, die sich aus der Formulierung des

Rubin Causal Model ableiten lässt: „Thus, at best the effect can only be estimated with confidence, but never measured with certainty.“ (Schmidt 1999:4).

Die Antwort auf dieses Problem mangelnder Beobachtbarkeit der kontrafaktischen Situation ist die Suche nach sogenannten Identifikationsstrategien. Darunter versteht man Annahmen (!), welche eine Konstruktion (!) einer kontrafaktischen Situation ermöglichen. Das bedeutet, man sucht nach verschiedenen Möglichkeiten, die unbeobachtbare Situation durch eine alternative beobachtbare Situation anzunähern. Ein grundlegendes Problem dabei ist die Tatsache, dass sich geförderte von nicht-geförderten Unternehmen in verschiedenen Merkmalen unterscheiden. Dieses sogenannte Sample-Selection Problem wird in diesem Unterkapitel ausführlich diskutiert. Vereinfacht gesprochen sollte ein Vergleich von geförderten Unternehmen mit nicht-geförderten Unternehmen keinen Vergleich von „Äpfeln mit Birnen“ darstellen. Entweder es werden tatsächlich vergleichbare Äpfel gefunden oder es muss eine Anpassung der Birnen an die Äpfel konstruiert (!) werden, so dass diese doch wieder mit den Äpfeln verglichen werden können.

Es gibt eine Reihe unterschiedlicher Identifikationsstrategien, die sich jeweils in ihren sogenannten Identifikationsannahmen unterscheiden. Diese Annahmen explizieren die Bedingungen unter denen die jeweilige Identifikationsstrategie eine valide Strategie zur Messung durchschnittlicher Maßnahmeneffekte auf die Maßnahmenteilnehmer (MTT) darstellt. Eine in diesem Zusammenhang weitere fundamentale Erkenntnis ist, dass sich die Frage, ob die getroffenen Annahmen zutreffen oder nicht einer formalen, statistischen Überprüfung entziehen. Deswegen ist a priori „keine Identifikationsannahme besser oder schlechter als eine andere“ (Bauer et al. 2009:146). Hier liegt auch der tiefere Grund, warum auch streng formalisierte, mathematisch-statistische Modelle auch völlig falsche Ergebnisse liefern können. Es ist dann aber nicht das Modell an sich falsch, sondern es sind die kritischen Annahmen nicht erfüllt. Daher sind ein Wissen und eine Sensibilität über die Bedeutung der getroffenen Annahmen für die Nutzer von ökonomischen Ergebnissen von entscheidender Bedeutung.

Im Folgenden werden zunächst die einfachsten und grundlegendsten Identifikationsstrategien dargestellt. Diese sind der Querschnittsvergleich, der Vorher-Nachher-Vergleich, und der Differenz-von-Differenzen-Ansatz. Schließlich wird auch die in Österreich bei FTI-Evaluierungen übliche Methode der direkten Frage nach kausalen Effekten beim Maßnahmenteilnehmer erörtert.

Im Kontext des hier aufgespannten konzeptionellen Denkrahmens wird deutlich, welche grundlegenden Probleme und letztlich mangelnde methodische Fundierung hier vorliegt.

Die Erörterung der Identifikationsstrategien und ihrer jeweiligen Probleme sollte klar machen, dass es letztlich komplexerer Methoden bedarf, um das Evaluationsproblem adäquat zu adressieren. Diese werden exemplarisch im darauf folgenden Abschnitt 5.2.4 dargestellt.

Querschnittsvergleich und Sample Selection Bias

Die erste hier zu besprechende Identifikationsstrategie ist der Querschnittsvergleich. Dieser beruht auf der Idee, dass die kontrafaktische Situation durch die Ergebnisgröße der Nichtteilnehmer ersetzt werden kann, um den MTT zu schätzen. Die Identifikationsannahme lautet also, dass sich

der Erwartungswert der Ergebnisgröße der Teilnehmer bei Nichtteilnahme genauso entwickelt hätte wie der Erwartungswert der Nichtteilnehmer, d.h.

$$E(Y_i(0)|D_i = 1) = E(Y_i(0)|D_i = 0). \quad (4)$$

Damit diese Identifikationsannahme eine unverzerrte Schätzung des durchschnittlichen Maßnahmeneffekts ermöglicht, muss gelten, das

$$E(Y_i(0)|D_i = 1) - E(Y_i(0)|D_i = 0) = 0. \quad (5)$$

Wenn diese Annahme zutrifft, haben wir einen beobachtbaren Ersatz für die kontrafaktische Situation gefunden, der eine valide Schätzung des MTT ermöglicht. Der Maßnahmeneffekt in Form ergibt sich dann als Querschnittsschätzer nach

$$MTT_Q = E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 0). \quad (6)$$

Alle Terme in Gleichung (6) sind beobachtbar und identifiziert. Das Evaluationsproblem ist gelöst, vorausgesetzt die Identifikationsannahme trifft zu. Von den Daten her verlangt der Querschnittsschätzer Beobachtungen über Teilnehmer und Nichtteilnehmer nach der Maßnahme. Welcher Zeitraum zwischen Maßnahme und Evaluation verstrichen sein sollte ist in Abhängigkeit von den Programmzielen und den damit zu erwartenden time-lags der Programmwirkung zu ermitteln. Liegt schließlich eine derartige Stichprobe vor, so kann Gleichung (6) durch einfache arithmetische Mittelwertbildung geschätzt werden:

$$MTT_Q = \frac{\sum_{i=1}^{N_1} Y_i(1)}{N_1} - \frac{\sum_{i=1}^{N_2} Y_i(0)}{N_2}, \quad (7)$$

wobei N_1 die Anzahl der Teilnehmer und N_0 die Anzahl der Nichtteilnehmer bezeichnen.

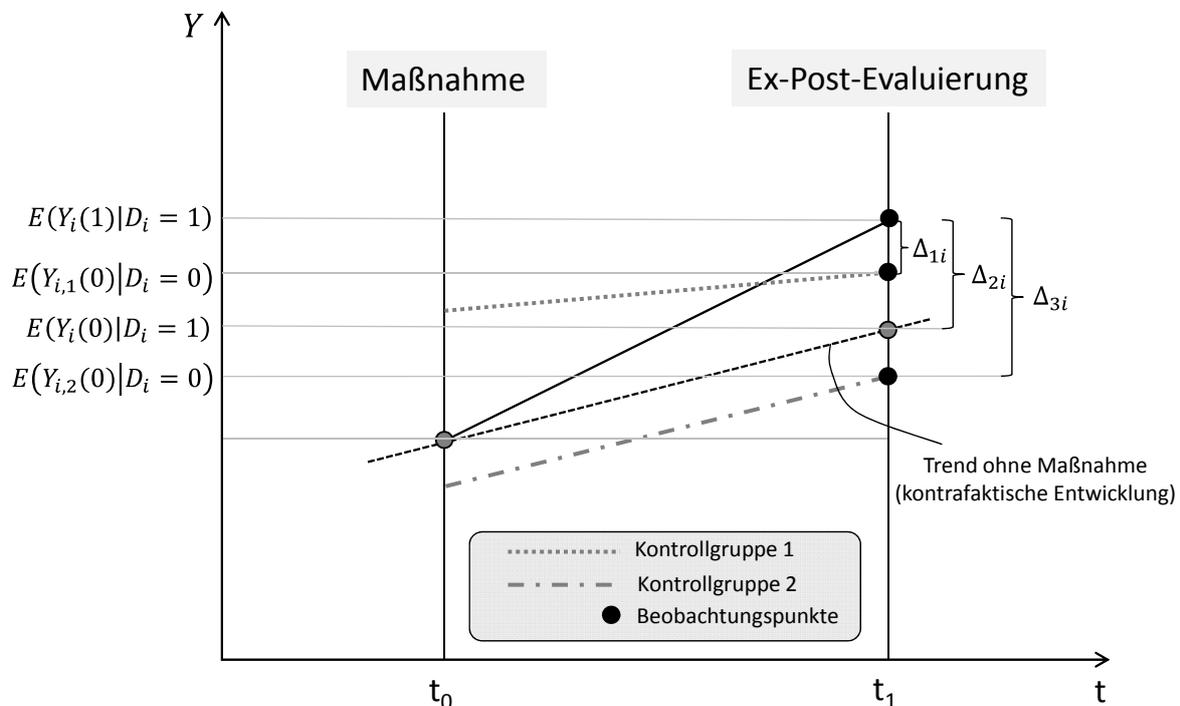
Welche Probleme können nun eine Verzerrung des Querschnittsvergleichs verursachen, also dazu führen, dass die Identifikationsannahme in Gleichung (4) nicht erfüllt ist? Abbildung 21 zeigt zunächst in grafischer Form die Problemlage auf. Dabei ist auf der Abszisse die Zeit und auf der Ordinate der Wert der Erfolgsgröße, d.h. die Patentzahl gemäß unserem Beispiel, aufgetragen. Die Förderung erfolgt dabei zum Zeitpunkt t_0 , wobei wieder der Einfachheit halber angenommen wird, dass die Förderung in der Ausbezahlung eines einmaligen Geldbetrags zur Durchführung eines Innovationsprojekts besteht. Nachdem eine bestimmte Zeit bis zur Patentierung verstreichen muss, erfolgt die Evaluierung erst zum Zeitpunkt t_1 . Weiterhin zeigt Abbildung 21 die realisierte Entwicklung für die Teilnehmer in Form der durchgezogenen schwarzen Linie und den nicht beobachtbaren, kontrafaktischen Trend, der die Entwicklung der Teilnehmer bei Nichtförderung beschreibt, durch eine gestrichelte Linie. Demnach ergibt sich der „wahre“ Maßnahmeneffekt im Sinne des MTT durch die Differenz Δ_{2i} . Nachdem aber die Situation $E(Y_i(0)|D_i=1)$ nicht beobachtbar ist, kann Δ_{2i} nicht direkt aus beobachtbaren Ergebnissen geschätzt werden. Deshalb ist in Abbildung 21 auch die Entwicklung zweier zur Verfügung stehender Kontrollgruppen eingezeichnet. Wie bei den Teilnehmern gilt auch hier, dass lediglich Beobachtungen zum Zeitpunkt t_1 vorliegen. Alle beobachteten Werte sind als schwarze Punkte auf der Vertikalen zum Zeitpunkt t_1 eingezeichnet.

Dabei zeigt sich, dass beide Kontrollgruppen die getroffene Identifikationsannahme des Querschnittsvergleichs verletzen. Das heißt es gilt

$$E(Y_i(0)|D_i = 1) \neq E(Y_{i,1}(0)|D_i = 0) \text{ und} \quad (8)$$

$$E(Y_i(0)|D_i = 1) \neq E(Y_{i,2}(0)|D_i = 0) \quad (9)$$

Abbildung 21: Der Querschnittsvergleich und mögliche Kontrollgruppen



Eigene Darstellung

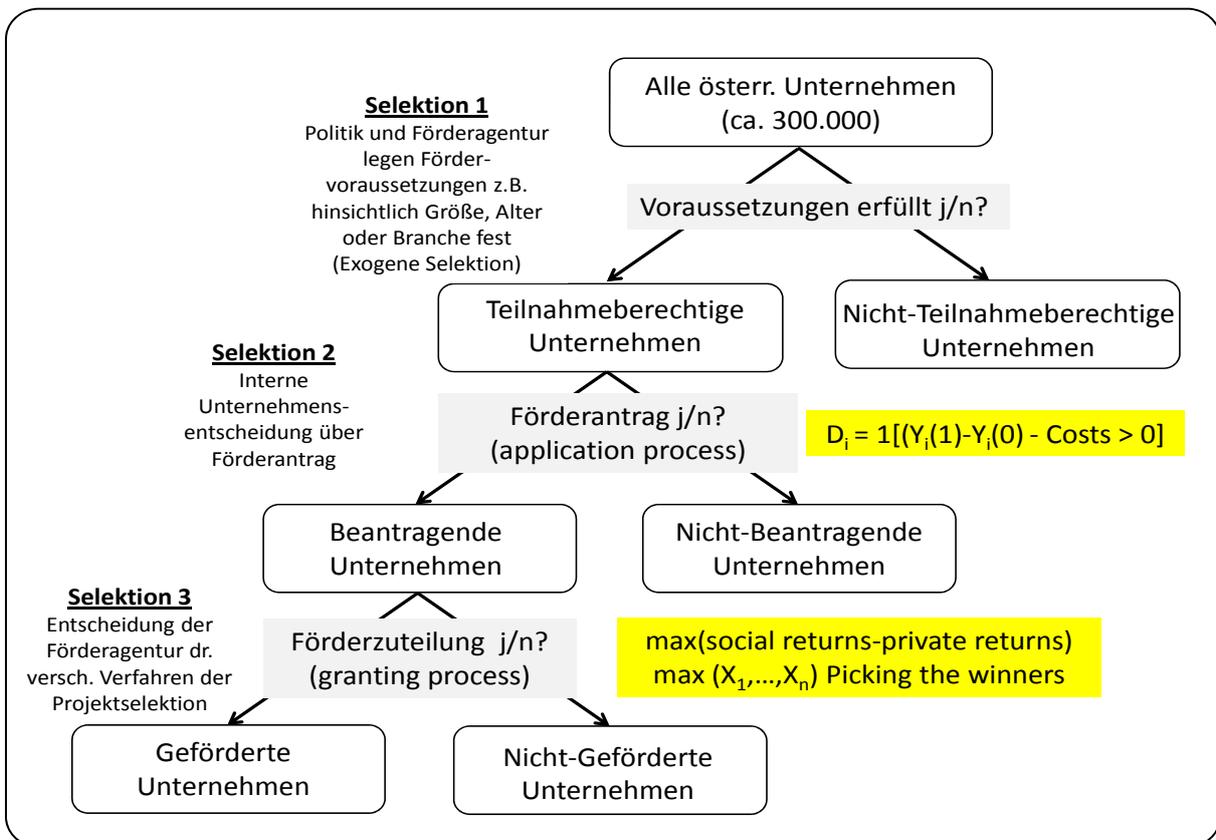
Die Verwendung von Kontrollgruppe 1, gekennzeichnet durch das Subskript 1, führt zu einer Unterschätzung des Maßnahmeneffekts, $\Delta_{1i} < \Delta_{2i}$. Das heißt die Teilnehmer hätten bei Nichtteilnahme eine schlechtere durchschnittliche Patentperformance realisiert als die Unternehmen von Kontrollgruppe 1. Umgekehrt verhält es sich bei Kontrollgruppe 2. Hier kommt es zu einer Überschätzung der Maßnahmeneffekte, $\Delta_{3i} > \Delta_{2i}$. Die Teilnehmer hätten sich auch bei Nichtteilnahme besser als die Unternehmen aus Kontrollgruppe 2 entwickelt. Es ist aus Abbildung 21 unmittelbar ersichtlich, dass eine Verletzung der Identifikationsannahmen schwerwiegende Fehler bei Evaluierungsergebnissen verursachen kann und die Konstruktion einer validen Kontrollgruppe von höchster Bedeutung ist. Hier sei nochmals betont, dass auch bei einer groben Verletzung der Identifikationsannahme im Prinzip keine Methode existiert, die auf eine derartige Verletzung testen könnte. Nachdem wir niemals wissen werden, welche Ergebnisse sich in der kontrafaktischen Situation eingestellt hätten, besteht eben keine Möglichkeit zur Überprüfung der Identifikationsannahme. Hier helfen nur substanzwissenschaftliche Argumentation und Sensibilität für die soziale bzw. ökonomische Realität.

Die Tatsache, dass die Nichtteilnehmer nicht einfach als Kontrollgruppe für die Teilnehmer verwendet werden können ist letztlich darin begründet, dass diese Gruppen systematisch

unterschiedliche Merkmale aufweisen. Wenn diese Merkmale auch einen Einfluss auf die Ergebnisgröße haben, liefert der Vergleich keinen validen Maßnahmenschätzer, weil dann unterscheiden sich die beiden Gruppen nicht mehr nur durch die Maßnahmenteilnahme, sondern durch weitere relevante Merkmale. Es werden also „Äpfel mit Birnen“ verglichen. Warum kommt es aber dazu, dass diese beiden Gruppen unterschiedlich sind? Die Antwort liegt im Selektionsverfahren der Maßnahmenteilnehmer begründet, das im Ergebnis bestimmt, welches Unternehmen eine Förderung erhält und welches nicht. Hier liegt ein Spezifikum von innovationspolitischen Evaluierungen relativ zu arbeitsmarktpolitischen Evaluierungen. Der Selektionsprozess ist generell durch einen höheren Grad an Komplexität gekennzeichnet und die Selektion durch die Förderagenturen wird als wichtige Leistung für das Innovationssystem aufgefasst.

Abbildung 21 zeigt einen idealtypischen Verlauf eines solchen Selektionsprozesses. Demnach gibt es 3 Stufen der Selektion. In einem ersten Schritt erfolgt eine Selektion nach Kriterien die anzeigen, ob ein Unternehmen überhaupt einen Antrag stellen kann oder nicht, z.B. sind einige Programme nach Branche oder Unternehmensgröße differenziert, d.h. es können etwa nur KMUs einen Antrag stellen. Die zweite Selektionsstufe betrifft die unternehmensinterne Entscheidung bei den teilnahmeberechtigten Unternehmen. Die Unternehmen entscheiden anhand eines Kosten-Nutzen Kalküls, ob sie einen Förderantrag stellen oder nicht.

Abbildung 22: Idealtypischer Selektionsprozess bei F&E-Programmen



Eigene Darstellung

Abbildung 21 enthält hierzu eine entsprechende Indikatorfunktion, die den Wert 1 annimmt, wenn die Bedingung in der Klammer erfüllt ist. Diese besagt, dass eine Teilnahme dann erfolgt, wenn die Vorteile der Teilnahme die Kosten der Teilnahme überwiegen. In einer dritten Selektionsphase schließlich erfolgt die Beurteilung der Anträge der Teilnehmer auf Förderung durch die Förderagentur, etwa im Rahmen einer internen Evaluierung oder durch Peer Review. Dabei können zwei Strategien der Agenturen unterschieden werden. Aus volkswirtschaftlicher Perspektive sollten jene Projekte gefördert werden, bei denen die Differenz zwischen privatem und sozialem Ertrag maximal wird. Weil aber die Informationsbasis für ein solches Entscheidungskalkül typischerweise nicht vorliegt, folgen die Agenturen einer „Picking the Winner“ Strategie. Dabei wird anhand einiger Unternehmensmerkmale X beurteilt, ob das Unternehmen gefördert werden soll oder nicht. Idealtypisch könnte man hier zwischen Merkmalen unterscheiden, welche einerseits das zur Förderung beantragte Projekt betreffen und andererseits Kategorien, welche das Unternehmen beschreiben. Typischerweise kommt es im Ergebnis dazu, dass Unternehmen mit erfolgreichen F&E-Aktivitäten in der Vergangenheit, mit einer eigenen F&E-Abteilung sowie größere und international aktive Unternehmen eine höhere Förderwahrscheinlichkeit aufweisen. Agenturen versuchen demnach jene Unternehmen zur fördern, die die höchste Erfolgswahrscheinlichkeit aufweisen, aus der Förderung z.B. ein Patent zu generieren oder eine andere Ergebnisgröße zu maximieren. Die innovationsökonomische Literatur hat dieses Ergebnis in zahlreichen empirischen Studien bestätigt (Clarysse/ Knockaert 2009, Canter 2011). Diese Strategie ist keineswegs unumstritten, wie Einiö (2009:1) argumentiert: „A major concern is that program managers may be encouraged to support projects with the best technical merits and the highest potential for commercial success. As these projects typically have high private returns they will be undertaken even in the absence of the support. In this case government support may induce only little additional R&D if any at all.“

Die Tatsache, dass der Selektionsprozess systematische Unterschiede zwischen Teilnehmern und Nichtteilnehmern produziert und diese Prozesse durch Akteure vorangetrieben werden, die jeweils unterschiedliche Zielfunktionen maximieren, ist *die* zentrale Erkenntnis der mikroökonomischen Evaluationsforschung schlechthin. Dieser Selektionsprozess erzeugt dadurch einen sogenannten „Sample Selection Bias“, was vereinfacht gesagt bedeutet, dass die Beobachtungen keine Zufallsstichprobe darstellen. Nachdem aber alle Standardverfahren der Ökonometrie auf der Annahme einer Zufallsstichprobe beruhen, braucht es die Anwendung komplexerer ökonomischer Verfahren, welche in der Lage sind, dieses Problem zu berücksichtigen.

Zusammengenommen zeigt Abbildung 21, dass der Selektionsprozess bei FTI-Projekten typischerweise von hoher Komplexität gekennzeichnet ist. Unterschiedliche Entscheidungskalküle multipler Akteure produzieren eine systematische Differenz zwischen geförderten und nicht-geförderten Unternehmen. Das Resultat ist gemäß der „Picking-the-Winner“ Strategie eine Positivselektion der geförderten Unternehmen im Vergleich zu den nicht-geförderten Unternehmen. Diese Situation korrespondiert mit dem überschätzten Maßnahmeneffekt Δ_{3i} in Abbildung 21. Damit stehen grundsätzlich alle Querschnittsvergleiche unter Verdacht, den Effekt der Maßnahme auf systematisch bedingte Performanceunterschiede auszudehnen, die nicht durch

die Förderung verursacht worden sind. Ein Beispiel hierfür findet sich in Box 14. Auch hier stellt sich das Problem von FTI-Evaluierungen anders dar als dies bei der Programmevaluierung von Maßnahmen der aktiven Arbeitsmarktpolitik der Fall ist. Nachdem diese Maßnahmen typischerweise Langzeitarbeitslose und andere Problemgruppen am Arbeitsmarkt adressieren, kommt es durch die Programmselektion zu Negativselektion relativ zur Vergleichsgruppe der Nichtteilnehmer. Demnach werden die Programmeffekte unterschätzt, weil die Nichtteilnehmer eine bessere Performance haben als die Teilnehmer bei Nichtteilnahme. Dies korrespondiert mit dem Maßnahmenschätzer Δ_{2i} in Abbildung 21.

Es sollte aus dieser Überlegung heraus klar werden, dass ein einfacher Vergleich zwischen geförderten und nicht-geförderten Unternehmen aufgrund des „Sample-Selection-Bias“ zu keiner validen Antwort auf die evaluatorische Fragestellung nach den kausalen Programmeffekten führen kann.

Box 14: Sample Selection Probleme bei der Evaluation spanischer FTI-Subventionen

Gonzalez und Pazo (2008) evaluieren in ihrer Studie die Inputadditionalität (zusätzlicher Innovationsinput aufgrund politischer Intervention) spanischer F&E-Subventionen anhand eines Matched-Pairs Ansatzes. Die verwendeten Daten sind ein Panel von Mikrodaten zu Industrieunternehmen für den Zeitraum von 1990-1999. In Summe handelt es sich um 9.455 Beobachtungen zu 2.214 Unternehmen. Zu Beginn ihrer Studie geben die Autoren zunächst einen deskriptiven statistischen Überblick über den Datensatz. Tabelle 16 zeigt dabei einen Vergleich der F&E-Intensitäten zwischen geförderten und nicht geförderten Unternehmen nach Unternehmensgröße differenziert. Für die geförderten Unternehmen ist zusätzlich eine Differenzierung in „total efforts“ und „private efforts“ angegeben. Die Differenz zwischen den beiden Größen ist die öffentliche Förderung.

Tabelle 16: F&E-Intensität bei geförderten und nicht geförderten Unternehmen im deskriptiven Vergleich

	With subsidies		Without subsidies
	Total effort	Private effort	Effort
Firms size			
≤20 workers	5.52	3.49	2.18
20–50 workers	3.42	2.55	1.73
50–100 workers	5.55	4.04	1.80
100–200 workers	3.41	2.64	1.89
200–500 workers	4.12	3.56	1.62
>500 workers	3.39	3.01	1.71
All firms	3.91	3.25	1.76

Gonzalez/Pazo 2008

Vergleicht man die private F&E-Intensität der geförderten Unternehmen mit jener der nicht-geförderten, so zeigt sich, dass die geförderten Unternehmen 3,25% ihres Umsatzes für F&E aufwenden, während dieser Wert bei nicht-geförderten bei 1,76% liegt. Die einfache Identifikationsannahme eines Querschnittsvergleichs würde diese Differenz als kausalen Fördereffekt interpretieren. Wie jedoch Gonzalez und Pazo (2008:375) ganz im Einklang mit der bisherigen Diskussion anmerken, ist dieser Vergleich problematisch: „This [i.e. der höhere F&E-Input bei geförderten Unternehmen], may be the consequence of the stimulating effect of subsidies or may simple be due to the subsidies being directed to firms that, even without subsidies, would make a higher-than-average technological effort. The objective of our paper is precisely to clarify this question.“

Gonzalez/ Pazo 2008

Welche Möglichkeiten bestehen, um im Rahmen eines Querschnittsvergleichs für einen Sample Selection Bias zu kontrollieren, sodass der Vergleich von Teilnehmern und Nichtteilnehmer letztlich doch eine valide Schätzung des MTT ermöglicht? Um das Problem zu lösen wäre grundsätzlich folgende Strategie zu wählen: In einem ersten Schritt werden jene Merkmale ermittelt, welche für

die Selektion in die Teilnehmer- oder Kontrollgruppe entscheidend waren. In einem zweiten Schritt wird genau für diese Merkmale kontrolliert.

Damit werden alle Unterschiede zwischen Teilnehmer- und Kontrollgruppe, die neben dem Merkmal Förderung bestehen, berücksichtigt und eine korrekte Identifikation der Maßnahmeneffekte möglich.

Angenommen der zur Verfügung stehende Datensatz enthält neben der Erfolgsgröße Y und der Teilnahmevariable D auch weitere exogene Variablen, d.h. beobachtbare Unternehmensmerkmale wie etwa Größe, Anzahl der F&E-Beschäftigten, Internationalisierungsgrad etc; diese sind in einem K -dimensionalen Vektor X_i zusammengefasst. Dann kann die ursprüngliche Identifikationsannahme verbessert werden, indem auch für systematische Unterschiede hinsichtlich dieser Unternehmensmerkmale zwischen Teilnehmern und Kontrollgruppe kontrolliert wird. Damit wird Identifikationsannahme (4) zu

$$E(Y_i(0)|X_i, D_i = 1) = E(Y_i(0)|X_i, D_i = 0) \quad (10)$$

Demnach ist die beobachtete Ergebnisgröße nicht nur konditional auf die Teilnahme, sondern auch konditional auf X . Jede beobachtbare Heterogenität zwischen den beiden Gruppen ist damit gleichsam „homogenisiert“. Wenn tatsächlich alle relevanten Unterschiede zwischen geförderten und nicht geförderten Unternehmen durch X erfasst werden, so stellt (10) eine valide Identifikationsstrategie dar. Aber auch damit sind noch nicht alle Zweifel beseitigt. Es könnte sein, dass neben einer Heterogenität aufgrund der beobachteten Merkmale X auch eine Heterogenität aufgrund von unbeobachtbaren Merkmalen besteht. Dies bezeichnet man entsprechend als unbeobachtbare Heterogenität. Diese Form der Heterogenität lässt sich per definitionem nicht nachweisen, sie lässt sich erneut nur aufgrund von Intuition und theoretischer Argumentation diskutieren.

Wenn diese unbeobachtbaren Merkmale, zusammengefasst in dem Vektor U , einerseits die Wahrscheinlichkeit der Maßnahmenteilnahme erhöhen und gleichzeitig einen positiven Effekt auf die Ergebnisgröße haben, stehen wir vor dem gleichen Problem wie zuvor. Liegt eine solche Konstellation vor, so kommt es erneut zur Überschätzung des Maßnahmeneffekts. Obwohl ein Teil der Unterschiede zwischen Teilnehmer und Kontrollgruppe durch die Merkmale U verursacht wurde, schlägt sie ein einfacher Regressionsansatz der Teilnahmevariablen D zu, die damit nach oben verzerrt ist.

Welche Merkmale könnten für eine unbeobachtete Heterogenität verantwortlich sein? Zunächst einmal kann es sein, dass technisch beobachtbare Merkmale, wie etwa der Standort eines Unternehmens oder der Gewinn aus verschiedenen Gründen nicht in dem zur Verfügung stehenden Datensatz enthalten sind. Daneben gibt es aber auch eine Gruppe von Variablen, die aufgrund der durch sie bezeichneten Eigenschaften nicht beobachtbar sind, oder wenn dann nur durch eine Reihe von – typischerweise nicht verfügbaren – Proxyvariablen angenähert werden können. Prominente Beispiele aus der Arbeitsmarktforschung betreffen etwa die Motivation oder die technische Begabung. Beide Variablen sind offensichtlich von Relevanz für Arbeitsmarktergebnisse, aber gleichzeitig de facto unbeobachtbar. Äquivalente Merkmale bei

Unternehmen wären etwa die Qualität des Managements und dessen Einbindung in politökonomische und unternehmerische Netzwerke, Risikoeinstellung, Qualität des jeweiligen Forschungsprojekts, technologische Absorptionsfähigkeit, etc.

Da auch diese Merkmale unbeobachtbar sind, aber sowohl die Förderwahrscheinlichkeit als auch die Ergebnisgröße positiv beeinflussen könnten, verletzt deren Nichtbeachtung die Identifikationsannahme des Querschnittsvergleichs. Eigentlich würden wir unseren Schätzer des Maßnahmeneffekts gerne auf nachstehender Identifikationsannahme basieren:

$$E(Y_i(0)|U_i, X_i, D_i = 1) = E(Y_i(0)|U_i, X_i, D_i = 0) \quad (11)$$

Leider ist das aber im Rahmen des Querschnittsvergleichs unmöglich, weil U eben per definitionem unbeobachtbar ist. Erneut stellt sich das Problem, dass wir diesen Sachverhalt nicht durch objektive, statistische Testverfahren überprüfen können. Ob eine solche unbeobachtete Heterogenität von Relevanz ist wird sicherlich auch von der Anzahl und Qualität der verfügbaren beobachtbaren Merkmale X abhängen. Aber auch die Programmcharakteristika und der jeweils programmspezifische Selektionsprozess spielen eine wichtige Rolle. Auf diese Problematik wird bei der Vorstellung von ökonometrischen Verfahren in Abschnitt 5.2.4 noch näher eingegangen.

Vorher-Nachher-Vergleich

Neben dem Querschnittsvergleich ist der Vorher-Nachher-Vergleich die zweite grundsätzlich mögliche Identifikationsstrategie. Die Teilnehmer dienen dabei als ihre eigene Kontrollgruppe. Hierfür werden nur Daten über die Teilnehmer benötigt, wobei diese für zumindest zwei Zeitpunkte vorliegen müssen. Wenn also keine Daten zu Nichtteilnehmern verfügbar sind, ist der Vorher-Nachher-Vergleich die einzig mögliche Identifikationsannahme. Die erste Beobachtung muss den Wert der Erfolgsgröße zeitlich vor der Maßnahme erfassen (t_0), die zweite Beobachtung sollte zeitlich nach Maßnahmenende stattfinden, wobei wieder der time-lag zwischen Förderung und Maßnahmeneffekte berücksichtigt werden muss. Daten, welche für die gleichen Unternehmen Beobachtungen zu mehreren Zeitpunkten enthalten, bezeichnet man als Paneldaten. Allerdings ist für einen Vorher-Nachher-Schätzer auch die Verwendung sogenannter gepoolter Querschnittsdaten möglich. Dabei handelt es sich um verschiedene Stichproben von z.B. geförderten Unternehmen, die zu unterschiedlichen Zeitpunkten erhoben wurden. Wenn es sich dabei um Zufallsstichproben handelt, ist das funktional äquivalent zur Situation mit echten Paneldaten. Die Tatsache, dass keine Daten über Nichtteilnehmer benötigt werden, ist ein potenzieller Vorteil dieses Schätzers, da Letztere bei Nichtverfügbarkeit von Sekundärdatenquellen teuer erhoben werden müssen. Die Identifikationsannahme beim Vorher-Nachher-Vergleich lautet:

$$E(Y_{i,t_1}(1)|X_i, D_i = 1) = E(Y_{i,t_0}(0)|X_i, D_i = 1) \quad (12)$$

In Worten: Die geförderten Unternehmen hätten sich bei Nichtförderung gegenüber ihrer Ausgangssituation nicht verändert, d.h. ihre Ergebnisgröße wäre konstant geblieben. Unmittelbar deutlich wird, dass im Vergleich zum Querschnittsvergleich das Problem unbeobachtbarer

Heterogenität zwischen Teilnahme und Kontrollgruppe beim Vorher-Nachher Vergleich obsolet wird, weil eine Identität zwischen den Mitgliedern dieser beiden Gruppen besteht.

Die mit dem Vorher-Nachher Schätzer verbundenen Probleme betreffen im Wesentlichen drei Aspekte. Erstens kann die Identifikationsannahme dadurch verletzt werden, dass eine exogene Entwicklung einen Einfluss auf die Ergebnisgröße ausübt. Dieser Einfluss wird fälschlicherweise dem Programm zugerechnet.

Ein prominentes Beispiel hierfür ist der Einfluss des Konjunkturzyklus, der z.B. den Absatz mit innovativen Produkten unabhängig von Innovationsförderungen in einem entscheidenden Maße mitbeeinflussen wird. Aus innovationsökonomischen Studien ist darüber hinaus bekannt, dass unternehmerische Innovationsinputs besonders akzentuierten Zyklen unterliegen. Damit wären Studien zur Inputadditionalität auf Basis eines Vorher-Nachher-Schätzers potenziell problematisch. Denkbar ist aber auch die Veränderung des regulatorischen Umfelds eines Unternehmens. So kam es etwa im Anschluss an das Kyoto-Protokoll zu einer starken Zunahme von Patenten im Bereich der Umwelttechnologien. Dies hat vor allem damit zu tun, dass die Unternehmen eine starke Expansion der betreffenden Märkte durch strengere Umweltschutzmaßnahmen erwarten. Ein Vorher-Nachher-Schätzer zur Evaluation einer umwelttechnologischen F&E-Förderung hätte diesen regulatorischen Einfluss Effekt dem betreffenden Programm zugeschrieben. Eine mögliche Lösung dieses Problems anhand eines Anwendungsbeispiels aus Österreich zeigt Box 15.

Box 15: Ökonometrische Evaluation der FFG-Förderungen anhand eines Vorher-Nachher-Schätzers

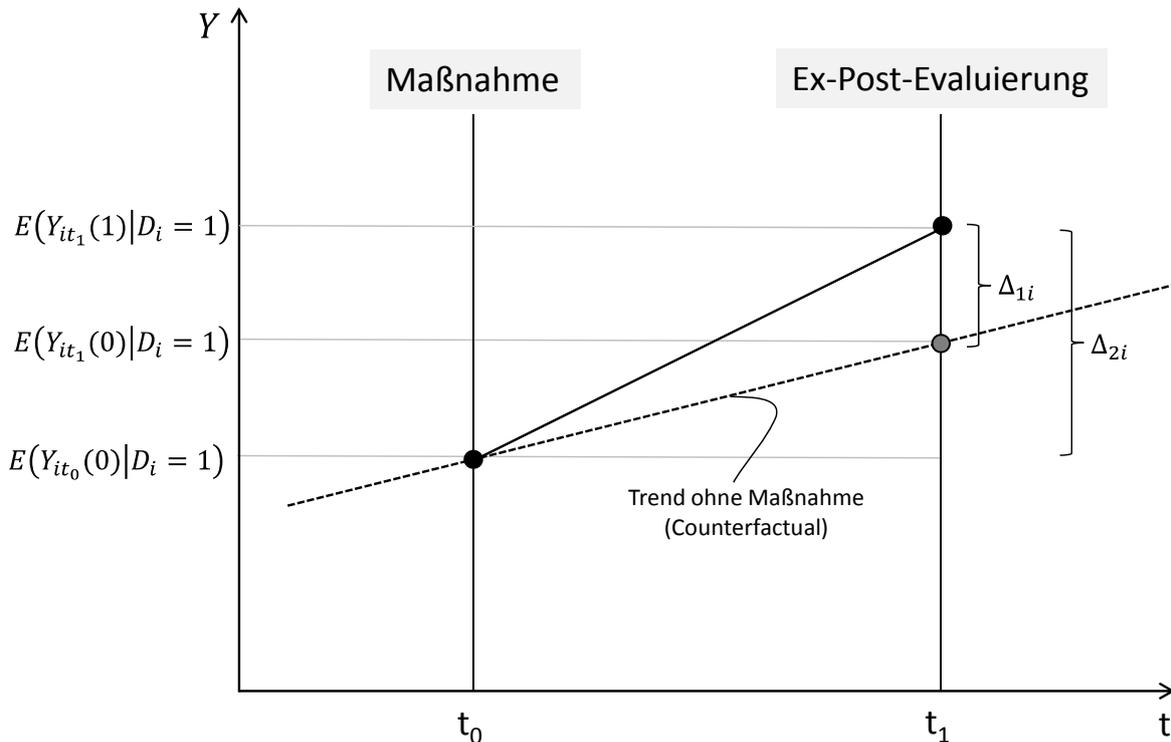
Streicher (2007) benutzte als Datengrundlage für die ökonometrische Evaluation der FFG-Förderungen (vormals FFF) die Förderdatenbank der FFG, um einen Datensatz von 342 Unternehmen für die Jahre 1997-2005 zu konstruieren. Mittels eines ökonometrischen Modells für Paneldaten (Fixed-Effects-Modell) wird schließlich eine Inputadditionalität bezüglich F&E-Ausgaben der Unternehmen geschätzt. Nachdem die Fördervariable metrisches Skalenniveau hat, kann der Fördereffekt von konjunkturellen Effekten isoliert werden, die mittels eines Perioden-Dummies berücksichtigt werden. Allerdings ist bei der Fördervariablen aus methodischen Gründen eine möglichst hohe Variation über die Jahre hinweg notwendig. Weiterhin wird diese Variable in der Studie aufgrund fehlender Daten fehlerhaft gemessen, weil nur die Finanzierung durch die FFG erfasst wurde und nicht jene von anderen Förderagenturen. Wenn Unternehmen, die von der FFG Förderungen erhalten haben, auch typischerweise von anderen Agenturen eher Förderungen bekommen, so führt dies zu einer Überschätzung der kausalen Effekte der FFG-Förderungen. Es wäre demnach – sowie aus vielen anderen Gründen – sinnvoll, eine möglichst zentrale Erfassung aller bedeutsamen F&E-Förderungen vorzunehmen. Zumindest sollten alle größeren Förderagenturen in einer gemeinsamen Datenbank mit ihren Förderungen auf Ebene der einzelnen Unternehmen erfasst werden (siehe Kapitel 7).

In der Studie wird folgende Identifikationsannahme getroffen: „This model allows for every firm to act, in a way, as its own control firm, in effect providing information on the firm’s behavior vis-à-vis different levels of support. This allows to overcome the major problem of the data base, the almost complete absence of firms which have some R&D activities, but which did not get any subsidy.“ Streicher (2007:11).

Das zweite grundlegende Problem beim Vorher-Nachher-Vergleich wird in der Literatur als „Ashenfelter Dip“ bezeichnet. Dieses nach dem Arbeitsmarktökonom Ashenfelter benannte Phänomen bezeichnet die oftmals beobachtete empirische Regularität, dass ökonomische Akteure in Erwartung einer Förderung ihren eigenen Ressourceninput reduzieren und damit eine Verschlechterung der jeweiligen Ergebnisgröße im zeitlichen Vorfeld der Förderung aufweisen. Erfolgt aber genau in diesem Zeitraum die Erfassung der Daten, kommt es zu einer Überschätzung des Fördereffekts. Es könnte z.B. sein, dass Unternehmen eines Technologiefeldes, die in der Zukunft eine bestimmte Förderung erwarten, ihren eigenen Faktorinput reduzieren und erst wieder bei Eintreten der Förderung diesen wieder auf normales Niveau anpassen. Drittens ist die Identifikationsannahme, dass ein Unternehmen bei Nichtförderung z.B. zwei Jahre lang keine Veränderung bei relevanten F&E-bezogenen Ergebnisgrößen, selbst bei konstantem makroökonomischem Umfeld, erfahren würde, zu hinterfragen. So könnten Unternehmen einfach die Suche nach anderen Finanzierungsquellen intensivieren oder Ressourcen innerhalb des Unternehmens umschichten. „Die meisten Unternehmen verfallen bei ausbleibendem Erfolg [oder ausbleibender Förderung] keineswegs in hilflose Lethargie, sondern reagieren mit eigenen Anstrengungen, um ihre Lage zu verbessern. Nur wenn der staatliche Eingriff mehr erreicht als die Unternehmen aus eigener Kraft geschafft hätten, kann man ernsthaft von einer Wirkung des staatlichen Eingriffs sprechen.“ (Sachverständigenrat Jahresgutachten 2010:229).

Abbildung 23 und Abbildung 24 stellen die damit einhergehenden Fehleinschätzungen der Maßnahmeneffekte graphisch dar. Abbildung 23 zeigt dabei zunächst eine Situation, bei der es zu einer deutlichen Überschätzung des Maßnahmeneffekts kommt. Während der MTT durch Δ_{1i} gegeben ist, identifiziert der Vorher-Nachher-Schätzer den kausalen Effekt Δ_{2i} . Das Unternehmen hätte auch ohne Förderung eine Verbesserung seiner Ergebnisgröße relativ zur Ausgangssituation zum Zeitpunkt vor der Förderung erfahren, etwa weil die Konjunkturlage vom Aufschwung in die Boom-Phase gedreht hat.

Abbildung 23: Vorher-Nachher-Schätzer bei positivem kontrafaktischem Trend

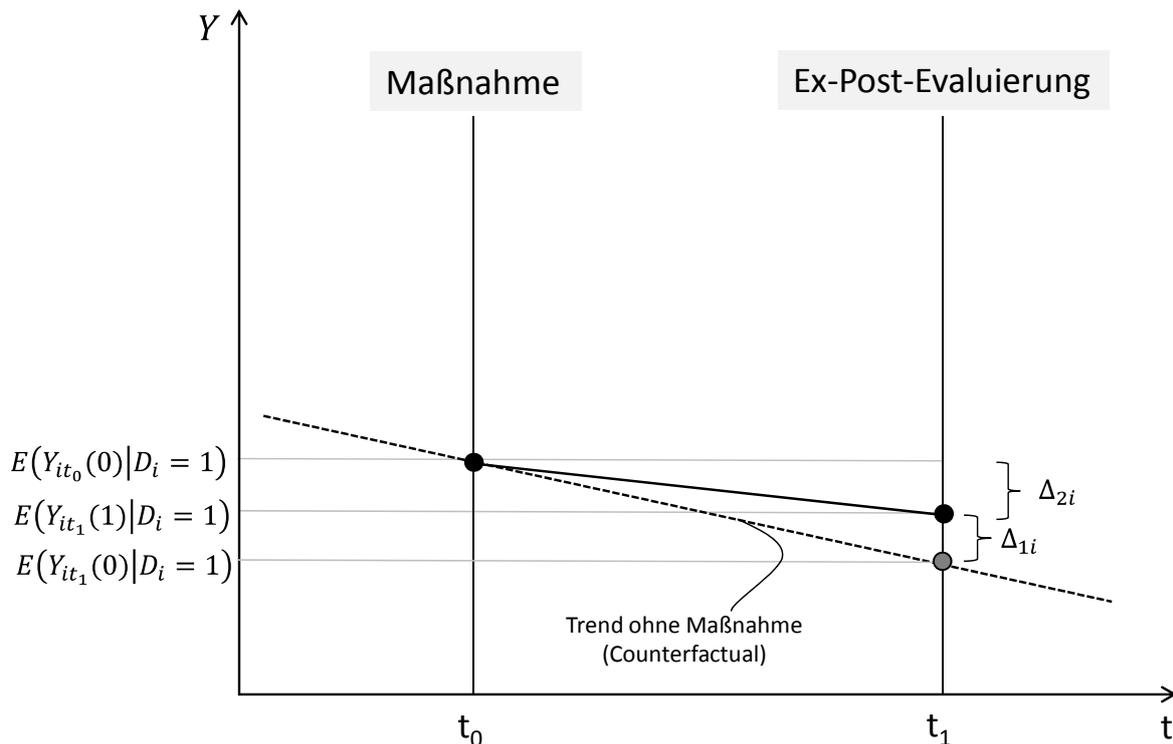


Eigene Darstellung

Abbildung 24 zeigt eine komplexere, weil kontraintuitive Konstellation, die jedoch keineswegs unrealistisch ist. Die beiden Beobachtungszeitpunkte liefern demnach einen deutlich negativen Maßnahmeneffekt in der Größenordnung von Δ_{2i} .

Nachdem etwa aufgrund eines konjunkturellen Einbruchs der Trend ohne Förderung zu einer noch schlechteren Performance geführt hätte, ist der wahre kausale Maßnahmeneffekt positiv und zwar in der Höhe von Δ_{1i} . Möglicherweise trifft diese Konstellation für einige Unternehmen in der aktuellen Krise zu.

Abbildung 24: Vorher-Nachher-Schätzer bei negativem kontrafaktischem Trend



Eigene Darstellung

Aufgrund dieser Probleme kommen Czarnitzki et al. (2003) zum Schluss, dass der Vorher-Nachher-Schätzer eher ungeeignet für FTI-Evaluierungen ist. Trotzdem können auch Zeitreihenbetrachtungen wertvolle Hinweise für Evaluierungen liefern, vor allem wenn die Zeitreihe einen längeren Zeitraum umfasst. Aber das Problem, inwieweit der Zeitreihenverlauf durch die Politik beeinflusst wurde, bleibt bestehen.

Differenz-von-Differenzen-Ansatz

Die beiden bisher diskutierten Identifikationsannahmen hatten jeweils spezifische Schwachstellen. Während die Identifikationsannahme des Querschnittsvergleichs durch unbeobachtete Heterogenität gefährdet ist, könnte der Vorher-Nachher-Vergleich durch zeitliche Effekte wie Konjunkturzyklen oder Änderungen im regulatorischen Umfeld verletzt werden. Beiden Probleme lassen sich prinzipiell durch eine Kombination dieser beiden Identifikationsannahmen lösen. Diese Idee liegt dem Differenz-von-Differenzen-Ansatz zugrunde. Von der Datenlage ist dieser Ansatz besonders anspruchsvoll, denn es werden Daten für Teilnehmer und Kontrollgruppe vor und nach der Maßnahme benötigt. Generell ist diese Identifikationsannahme eher selten in der innovationsökonomischen Literatur anzutreffen. Eine Ausnahme wurde in

Box 12 dargestellt.

Abbildung 25 zeigt die entsprechende Datenkonstellation.

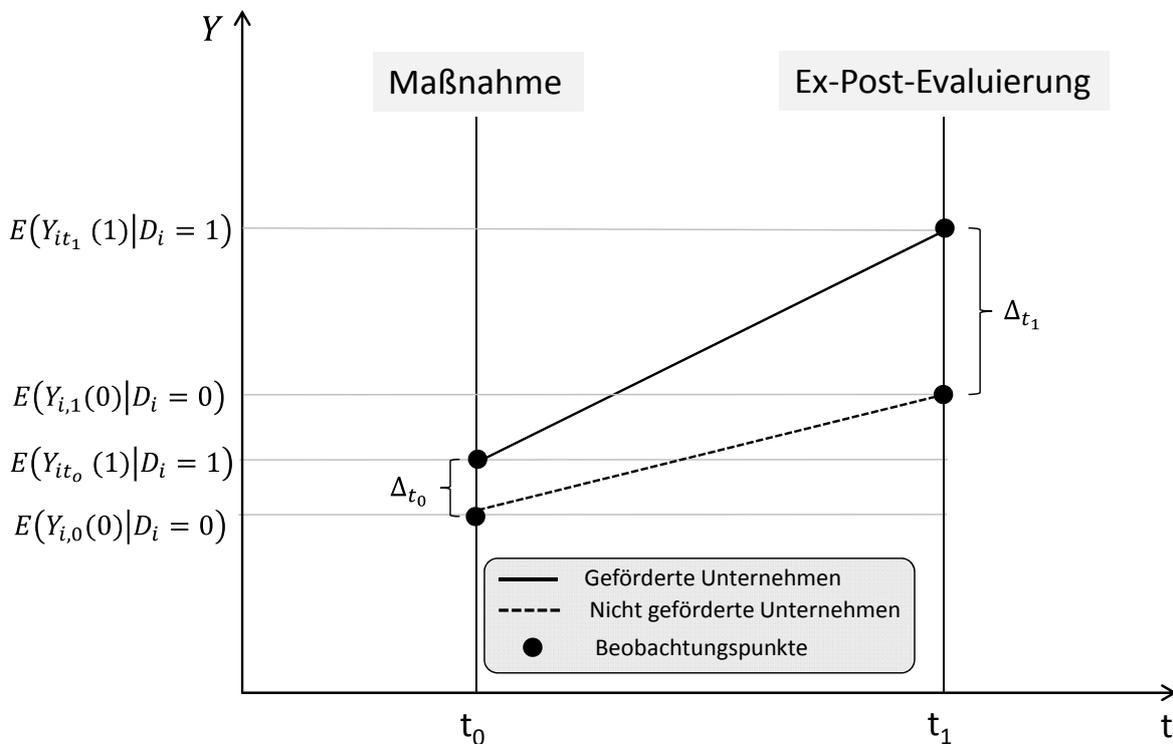
Es liegen demnach sowohl Daten über die Entwicklung der Teilnehmer als auch über die Entwicklung der Ergebnisgröße bei den Nichtteilnehmern vor. Dabei dienen die Nichtteilnehmer als Kontrollgruppe für die Teilnehmer, und die Veränderung der Erfolgsgröße zwischen den beiden Zeitpunkten t_0 und t_1 wird zur Schätzung des Maßnahmeneffekts herangezogen. Genauer gesagt wird eine Differenz von Differenzen gebildet, die auf folgender Identifikationsannahme beruht:¹⁴

$$E(Y_{i,t_1}(1)|D_i = 1) = E(Y_{it_1}(0)|D_i = 1) = \quad (13)$$

$$\{E(Y_{i,t_1}(1)|D_i = 1) - E(Y_{it_1}(0)|D_i = 0)\} -$$

$$\{E(Y_{i,t_0}(1)|D_i = 1) - E(Y_{it_0}(0)|D_i = 0)\}.$$

Abbildung 25: Der Differenz-von-Differenzen-Ansatz



Eigene Darstellung

Es wird also angenommen, dass die Differenz zwischen den Teilnehmern und Nichtteilnehmern über die Zeit hinweg konstant geblieben wäre, hätten die Teilnehmer nicht an der Maßnahme teilgenommen. Damit wird auch für Differenzen in Form unbeobachteter Heterogenität zwischen Teilnahme- und Kontrollgruppe kontrolliert, weil der dadurch verursachte Performanceunterschied durch die Differenzenbildung vor der Maßnahme berücksichtigt wird. Auch Konjunkturzyklen verursachen keine Probleme bei der Schätzung des Maßnahmeneffekts, vorausgesetzt beide Gruppen erfahren dadurch den gleichen Effekt. Sollte dies nicht der Fall sein, so gibt auch der Differenz-von-Differenzen-Schätzer nicht den kausalen Maßnahmeneffekt wieder.

¹⁴ Der Einfachheit halber wurden hier die exogenen Variablen X weggelassen.

Die direkte kausale Frage als Identifikationsstrategie

Alle bislang vorgestellten Identifikationsannahmen stellen die Ausnahme bei österreichischen Evaluierungen dar. In den allermeisten Fällen kommt es zu einer Konstruktion der kontrafaktischen Situation über die direkte Frage bei den teilnehmenden Unternehmen. Ein typisches Beispiel hierfür ist das FFG-Wirkungsmonitoring (siehe Box 16). Aber auch andere Evaluierungen, wie etwa die FFF-Evaluierung von 2004 basierten u.a. auf Befragungen mit kontrafaktischen Fragen an die geförderten Unternehmen (Falk 2006).

Box 16: Die Identifikationsannahmen des FFG-Wirkungsmonitorings

Das heutige FFG-Wirkungsmonitoring wird seit über 30 Jahren von der KMU-Forschung Austria durchgeführt. 2010 wurde das ursprünglich nur auf die Basisprogramme beschränkte Monitoring auf die thematischen Programme der FFG ausgedehnt. Dies steht auch im Zusammenhang mit der Tatsache, dass einige Kennzahlen des Wirkungsmonitorings nunmehr auch als Indikatoren im Rahmen der Wirkungsorientierten Haushaltsführung herangezogen werden. Damit kommt es zu einer deutlichen Aufwertung des FFG Wirkungsmonitorings. Nicht zuletzt werden die Ergebnisse auch in der Öffentlichkeit zur Legitimation der Förderaktivitäten der FFG benutzt. Es erscheint daher sinnvoll, sich mit den methodischen Voraussetzungen des Wirkungsmonitorings zu befassen. Die dort getroffene Identifikationsannahme entspricht der direkten kausalen Frage an die Förderempfänger. Am Ende des Fragebogens von 2010 findet sich die hierfür entscheidende Fragestellung (Abbildung 26). Das Ergebnis für die Befragung von 2010 zeigt Tabelle 17. Die damit verbundenen Identifikationsannahmen sind im Fließtext erörtert. Nachdem beide Identifikationsannahmen einigermaßen restriktiv sind und noch dazu voneinander abhängen, sind methodische Zweifel an dieser Vorgehensweise angebracht. Bedenkt man, dass die Befragung vier Jahre nach der Förderung stattgefunden hat (um time-lags in der Förderwirkung zu berücksichtigen) und der Anteil öffentlicher Subventionen an den gesamten F&E-Ausgaben des Unternehmenssektors ca. 1/10 ausmacht, so wird die Herausforderung bei der Beantwortung dieser Frage zusätzlich deutlich.

Abbildung 26: Die kontrafaktische Frage im FFG-Wirkungsmonitoring 2010

27. Hätten Sie das Projekt auch **ohne Förderung** durchgeführt?

ja, unverändert

ja, aber mit Zeitverzögerung

ja, in überwiegendem Ausmaß

ja, aber in deutlich geringerem Ausmaß

ja, aber ohne Kooperationspartner

ja, aber mit anderen Kooperationspartnern

ja, aber mit niedrigerem inhaltlichem / technischem Anspruch

nein, weil.....

KMU-Forschung Austria 2011

Tabelle 17: Antworten auf die kontrafaktische Frage 2006

	in %
ja, unverändert	5%
ja, in überwiegendem Ausmaß	8%
ja, aber ohne Kooperationspartner	1%
ja, aber mit niedrigerem technischen Anspruch	16%
ja, aber mit Zeitverzögerung	28%
ja, aber in deutlich geringerem Ausmaß	37%
ja, aber mit anderen Kooperationspartnern	1%
nein	22%
Quelle: KMU FORSCHUNG AUSTRIA, N = 343	

KMU-Forschung Austria 2011

Es gilt allerdings noch ein weiteres Problem zu beachten: Dies sei am Beispiel des Fördermultiplikators exemplifiziert. Dieser ergibt sich als Quotient aus Lizenzerlösen und Zusatzumsätzen, dividiert durch den Barwert der Fördermittel. Im Bericht zum Wirkungsmonitoring wird dies als eine Input-output-Relation interpretiert. Die Vorstellung ist also, dass ein Förder-Euro Zusatzumsätze und Lizenzerlöse der Unternehmen in der Höhe von x Euro verursacht. Die zur Berechnung dieser Größe notwendigen Fragen betreffen unter anderem die Fragen 16-18 im Fragebogen (Abbildung 27). Analysiert man die Fragestellung, so mag man sich erneut die Frage stellen, ob das nicht Fragen sind, welche das Unternehmen grundsätzlich überfordern.

Bedeutsamer ist jedoch die spezifische Formulierung der Frage, weil hier nach den Umsatzwirkungen des Projekts insgesamt gefragt wird. Damit wird implizit angenommen, dass das Projekt ohne Förderung nicht realisiert worden wäre. Hier werden also genau jene aus methodischer Sicht schwerwiegenden Fehler gemacht, die in der Literatur auch als Verwechslung von Brutto- und Nettoeffekten bekannt sind; schwerwiegend deshalb, weil damit ein Missverständnis von notwendigen Identifikationsannahmen zur Evaluierung kausaler Programmeffekte vorliegt oder jedenfalls befördert wird. Bedenkt man, dass das Wirkungsmonitoring weiter an Bedeutung gewinnen soll, so sind dies denkbar schlechte Voraussetzungen für ein methodisch korrektes Monitoring von Politikeffekten: Es werden gerade nicht Politikeffekte, sondern eigentlich Projekteffekte ermittelt. Der methodisch korrekte Wert im Nenner der oben erwähnten Division zur Berechnung des Fördermultiplikators müsste demnach in der Summe der Ressourcen bestehen, die ein Unternehmen für das jeweilige Projekt aufgewendet hat – und eben nicht nur die F&E-Förderung. Welche sinnvolle Interpretation aus FTI-politischer Perspektive mit dieser Größe verbunden ist, ist jedoch keineswegs offensichtlich.

Abbildung 27: Fragen im FFG-Wirkungsmonitoring zur Ermittlung des Fördermultiplikators

- 16. Konnten** seit Abschluss des Projektes in Ihrem Unternehmen durch die Verwertung der (dem Projekt direkt zurechenbaren) Ergebnisse **Lizenz Erlöse** realisiert werden?
- ja, Lizenz Erlöse insgesamt in € davon Erlöse im Ausland in €
- nein Die Projektergebnisse sind nicht lizenzierbar
- 17. Konnten** in Ihrem Unternehmen seit Abschluss des Projektes durch die Verwertung der Projektergebnisse **zusätzliche**, dem Projekt direkt zurechenbare **Umsätze** realisiert werden?
- ja, Zusatzumsatz von insgesamt in € davon Exportumsatz in €
- ja, aber die Höhe ist nicht spezifizierbar
- nein
- 18. Wären** in Ihrem Unternehmen ohne den Projekterfolg Umsatzeinbußen in anderen Bereichen unvermeidlich gewesen; d. h. die Ergebnisse trugen dazu bei, dass **Umsätze** seit Abschluss des Projektes **gesichert** werden konnten?
- ja, gesicherte Umsätze insgesamt in € davon Exportumsatz in €
- ja, aber die Höhe ist nicht spezifizierbar
- nein

KMU-Forschung Austria 2011

Wie müsste eine korrekte Formulierung der Frage zur Ermittlung des kausalen Programmeffekts, etwa bei Frage 17, lauten? Die ursprüngliche Formulierung der Frage lautet:

- „Konnten in Ihrem Unternehmen, seit Abschluss des Projekts durch die Verwertung der Projektergebnisse zusätzliche, dem Projekt direkt zurechenbare Umsätze realisiert werden?“

Bei dieser Formulierung wird aber nach dem Projekteffekt gefragt, der eben sehr verschieden von dem Politikeffekt sein kann. Um Letzteren zu erfragen müsste die Frage vielmehr folgendermaßen lauten:

- „Konnten in Ihrem Unternehmen, seit Abschluss des Projekts durch die Verwertung der Projektergebnisse zusätzliche, der Projektförderung direkt zurechenbare Umsätze realisiert werden?“

Spätestens hier sollte jedoch vollkommen klar sein, dass eine solche Frage auch jeden noch so gewissenhaft antwortenden Manager überfordern muss. Die Antwort hierauf kann nur durch methodisch nachvollziehbare Identifikationsannahmen geschätzt aber keinesfalls direkt abgefragt werden!

Diese eben exemplarisch geschilderte Problemlage wird noch größer weil bis auf die letzte von 27 Fragen alle Fragen ausnahmslos dieser falschen Logik der Ermittlung kausaler Maßnahmeneffekte folgen. Zusammenfassend kann festgestellt werden, dass das Wirkungsmonitoring wesentliche methodische Schwächen aufweist, die in deutlichem Kontrast zur Bedeutung desselben stehen.

Die grundlegende Konstellation bei dieser Identifikationsstrategie zeigt Abbildung 28. Es liegt nur eine einzige Beobachtung vor, welche die Erfolgsgröße nach Maßnahmenende angibt. Die kontrafaktische Situation bezeichnet einen Punkt auf der Vertikalen über dem Evaluierungszeitpunkt. Diese muss vom befragten Unternehmen bestimmt werden. Offensichtlich ist diese Identifikationsstrategie zunächst einmal vor allem durch einen relativen Datenmangel im Vergleich zu den anderen Identifikationsstrategien gekennzeichnet.

Es liegt genau genommen nur eine Punktbeobachtung vor, was erhebliche Probleme schafft: “In fact, with only a point observation after treatment, it is impossible to reach a conclusion about the impact.” (Khandker et al. 2010:22). Was liegt also näher als die Teilnehmer über ihre Situation bei Nichtförderung zu befragen? Dieses zunächst trivial anmutende Evaluationsverfahren ist jedoch keineswegs unproblematisch und die genauen Identifikationsannahmen sind sehr rigoros.

Abbildung 28: Direkte Frage nach der kontrafaktischen Situation als Identifikationsannahme



Eigene Darstellung

Welche Voraussetzungen, i.e. Identifikationsannahmen müssen erfüllt sein, damit eine solche Methode ein valide Abschätzung von Maßnahmeneffekten ermöglicht? Im Wesentlichen handelt es sich um zwei Identifikationsannahmen. Erstens muss angenommen werden, dass Unternehmen tatsächlich die kontrafaktische Situation identifizieren können. Dies ist die notwendige Identifikationsannahme.

Zweitens (vorausgesetzt die erste Identifikationsannahme ist erfüllt) muss angenommen werden, dass Unternehmen die Wahrheit bei der Befragung sagen und damit ihr Wissen über die kontrafaktische Situation ohne Verzerrung preisgeben. Erst wenn beide Identifikationsannahmen erfüllt sind, kann von einer hinreichenden Voraussetzung zur Identifikation kausaler Fördereffekte ausgegangen werden.

Die Erkenntnis, dass beide Identifikationsannahmen einander bedingen, schafft eine dritte Identifikationsannahme. Selbst wenn die Unternehmen ehrlich auf die Fragen antworten, könnten sie aufgrund fehlerhafter Einschätzungen über ihre kontrafaktische Situation eine ganz falsche Auskunft geben.

Beide Annahmen wurden in der Literatur stark kritisiert und generell muss festgestellt werden, dass die direkte Frage als Identifikationsstrategie ohne wissenschaftliche Fundierung und auch als Evaluationsmethode ungeeignet und höchst fragwürdig ist.

Die oftmalige Anwendung in österreichischen Evaluierungen steht in einem krassen Missverhältnis zur methodischen Kapazität dieser Identifikationsstrategie. Erstaunlich ist, dass so gut wie keine tiefergehenden methodischen Reflexionen existierten. So gibt es z.B. bei der Forschung zu öffentlichen Gütern umfangreiche Forschungen zur Frage, wie Individuen über ihre Zahlungsbereitschaft befragt werden können, so dass ihre Antwort möglichst ehrlich ausfällt. Die Gefahr ist offensichtlich: Nachdem das befragte Individuum nicht dafür zahlen muss, wird es unter normalen Umständen seine Zahlungsbereitschaft über- oder untertreiben, je nachdem ob das öffentliche Gut einen positiven oder negativen Einfluss auf den Nutzen des Individuums hat. "Care has to be taken to design questions in such ways as to minimize the risk that people don't respond truthfully. In recent years, cunning methods have been devised by economists to ensure that people don't exaggerate their fondness for those goods, especially in circumstances where they don't have to pay for them" (Dasgupta 2007: 144). Ähnliche methodische Reflexionen wären auch bei der direkten Frage als Identifikationsstrategie wichtig, weil die Gefahr, dass Unternehmen den Nutzen der Förderung übertreiben ebenfalls offensichtlich ist.

Welche konkreten Gründe könnten nun dazu führen, dass die Identifikationsannahmen verletzt werden?

Box 17 fasst diese anhand von Stellungnahmen prominenter Innovationsökonominnen zusammen. Interessant ist die Einschätzung, dass das notwendige Wissen auf Seiten der Unternehmen fehlt, um ihr eigenes Kontrafaktum zu bestimmen. Nachdem dies aber die Voraussetzung für die Validität dieser Identifikationsstrategie ist, hilft auch die Ehrlichkeit der Unternehmen nicht weiter, wenngleich eigentlich hier alles gegen ein ehrliches Antwortverhalten spricht. Freilich könnte man darauf vertrauen, dass auch Unternehmen an der Optimierung des öffentlichen Mitteleinsatzes interessiert sind und gleichsam auch eine soziale Wohlfahrtsfunktion in ihren Antworten beachten. (Aber wer möchte auf diese Annahme vertrauen?)

Interessant ist auch die zweite Stellungnahme in Box 17, die sich auf klassische Kosten-Nutzen Analysen bezieht. Bei diesen steht am Ende regelhaft eine Zahl, deren Zustandekommen aber auf „dubiosen“ Fragen basieren kann. Das sollte der kritische Nutzer solcher Analysen beachten und entsprechenden Wert auf eine methodisch anerkannte und explizite Identifikationsstrategie legen.

Box 17: Einwände gegen die direkte Frage als Identifikationsstrategie

“The third approach comes closest to the core of additionality, by asking the firm directly about the counterfactual – would it have done the work without assistance and if so how would it have been different. This presumes that the respondent is honest (usually they are) [Woher weiß Georghiou das? Er kann es einfach nicht wissen, weil die kontrafaktische Situation unbeobachtbar ist.] but more importantly that they are capable of performing the counterfactual analysis themselves.”

Georghiou (2004:64)

“In one of the bedrock papers in this field, Edwin Mansfield, the late University of Pennsylvania economist, estimated that academic research delivered an annual rate of return of 28% (Mansfield, *Research Policy*, 20, 1-12). The figure has been widely quoted ever since. But Mansfield reached this estimate by interviewing chief executives, asking them what proportion of their companies’ innovation was derived from university research and, in effect, demanding that they come up with a number [Herv. d. Verf.]. “He was asking an impossible question”, says Ben Martin, a former director of the Science and Technology Policy Research Unit at the University of Sussex, UK. “Methodologically, this was a dubious thing to do [Herv. d. Verf.]”

Macillwain (2010:683)

„In zahlreichen Befragungen werden daher hypothetische Fragen gestellt, wie z.B. „Hätten Sie das Projekt auch ohne staatliche Unterstützung durchgeführt?“ Derartige Fragen überfordern nicht nur den Befragten, sie lassen auch die Ergebnisse der Untersuchung zweifelhaft erscheinen, da diese auf Vermutungen basieren. Zudem, und dieses Argument ist noch gewichtiger, ist der InterviewpartnerInnen in einer Befragung geneigt, die für ihn positive Fördermaßnahme als ‚zwingend notwendig‘ einzustufen, weil er auch in Zukunft auf eine finanzielle Unterstützung hofft.“

Czarnitzki et al. (2003:45)

Zusammengenommen kann festgestellt werden, dass die direkte Frage als Identifikationsstrategie zwar sehr häufig bei FTI-Evaluierungen angewandt wird, aber die methodisch-wissenschaftliche Fundierung der damit verbundenen Identifikationsannahmen sehr dünn ist. Die Identifikationsannahmen sind verglichen zu jenen beim Querschnittsvergleich oder Vorher-Nachher Schätzer wesentlich problematischer. Überfordernde Informationsleistung und ein Vertrauen auf ehrliche Antworten auch gegen die eigenen Interessen der Gewinnmaximierung sind Identifikationsannahmen, deren argumentative Verteidigung wohl kaum möglich sein dürfte.

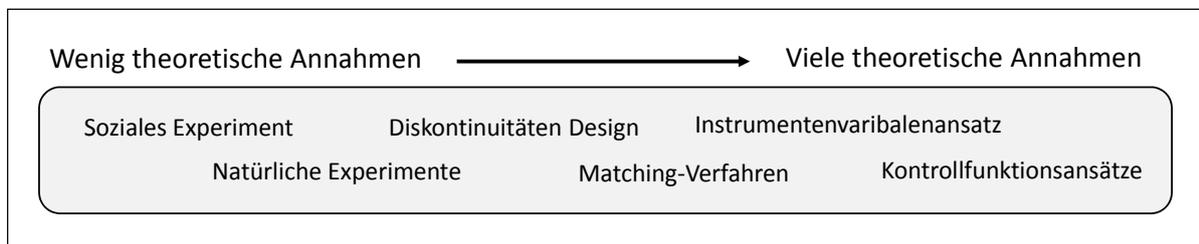
5.2.4 Ausgewählte mikroökonomische Methoden zur Programmevaluierung

Abschnitt 5.2.3 hat deutlich gemacht, dass einfache Identifikationsstrategien rasch an ihre Grenzen stoßen können. Allerdings sind diese Möglichkeiten zur Identifikation kausaler Effekte von Bedeutung um die Notwendigkeit für komplexere Methoden zu erkennen sowie um Evaluierungsergebnissen gegenüber eine kritische Haltung einnehmen zu können. Mikroökonomischen Methoden lassen sich anhand unterschiedlicher Merkmale klassifizieren.

Im Folgenden werden zwei derartige Einteilungen präsentiert, die wichtige Informationen über die Unterschiede zwischen den Methoden geben.

Zunächst einmal kann nach dem „Theoriegehalt“ einer Methode unterschieden werden (Blundell/ Costa Dias 2008). Abbildung 29 zeigt diesbezüglich das Methodenspektrum entlang eines Kontinuums zunehmender Theorieintensität. Auf der linken Seite dieses Spektrums befindet sich das soziale Experiment. Der Mechanismus der Randomisierung macht alle weiteren theoretischen Annahmen überflüssig. So ist etwa die Suche nach Kontrollvariablen bei einer hinreichend großen Fallzahl erst gar nicht notwendig. Auch ist kein theoretisches Wissen über die konkreten Wirkungsmechanismen notwendig. Diese müssen nicht modelliert werden. Allerdings muss dafür der Ablauf des Experiments einige wichtige Annahmen erfüllen, deren Einhaltung in der Evaluierungspraxis keineswegs selbstverständlich ist. Auf der anderen Seite des Spektrums finden sich Studien die Kontrollfunktionsansätze anwenden. Hierbei wird versucht, den Teilnahmeprozess als mikroökonomisches Entscheidungskalkül darzustellen. Eine explizite ökonometrische Modellierung von solchen Entscheidungssituationen und deren Interdependenzen ist Teil des sogenannten „Structural Econometric Approach“ (Cameron/ Trivedi 2005). Dabei müssen zahlreiche, aus der mikroökonomischen Theorie abgeleitete Annahmen getroffen werden. Es wird also wesentlich mehr Wissen „über die Welt“ vorausgesetzt, als dies bei einem Experiment der Fall ist. Die geringere Notwendigkeit zum Treffen zahlreicher, nicht testbarer Annahmen bei sozialen Experimenten macht diese zur glaubwürdigsten Methode für Kausalanalysen.

Abbildung 29: Klassifikation mikroökonomischer Methoden nach der Notwendigkeit zur theoriegeleiteten Modellierung



Eigene Darstellung

Eine weitere für die Methodenselektion wichtige Differenzierung erfolgt nach dem Selektionsmechanismus (Assignment Mechanism), der bei der angewandten Methodik vorliegen muss, damit diese valide Schätzungen der Maßnahmeneffekte ermöglicht. Tabelle 18 stellt die entsprechende Klassifikation der Methoden dar. Dabei wird zwischen drei unterschiedlichen Selektions- oder Zuweisungsprozessen differenziert. Der vorhergehende Abschnitt hat deutlich gemacht, dass dem Selektionsmechanismus überragende Bedeutung bei der Anwendung mikroökonomischer Methoden zukommt weil dieser den alles entscheidenden Selection Bias zwischen Teilnehmern und Nichtteilnehmern produziert. Entsprechend merken Blundell und Coast Dias (2008:2) an „Unless there is a convincing case for the reliability of the assignment mechanism being used, the results of the evaluation are unlikely to convince the thoughtful skeptic.“ Aufgrund dieser zentralen Rolle des Selektionsprozesses liegt einer der Schlüssel für eine Verbesserung von Evaluierungsergebnissen in einer Berücksichtigung der Evaluation bereits beim Selektionsprozess.

Genau diese Forderung stellt Jaffe (2002) und erwähnt z.B. die Randomisierung der Mittelvergabe als eine Möglichkeit in diesem Zusammenhang. Damit wäre einerseits ein Maximum an Glaubwürdigkeit bei der Evaluation möglich und andererseits könnten auch Mängel des Peer Review-Verfahrens reduziert werden. Bislang fehlt für diesen innovativen Ansatz der politökonomische „Rückenwind“. In den USA wird das soziale Experiment bereits regelmäßig in einigen Politikfeldern zur Politikevaluation angewandt. Ein Beispiel für eine FTI-Evaluation auf Basis einer randomisierten Selektion ist die Evaluation des niederländischen Innovationsschecks (Cornet et al. 2006).

Würde es dieses Sample-Selection-Problem nicht geben, wäre Evaluation sehr einfach. Während die randomisierte Zuweisung zur Teilnehmer- und Kontrollgruppe keiner weitergehenden Erörterung bedarf, ist die Unterscheidung von „Selction on observeables“ vs „Selection on unobservables“ komplexer und wurde bereits im Zusammenhang mit dem Querschnittsvergleich in Abschnitt 5.2.3 erörtert. Nachdem es sich hier um einen kritischen Punkt der Evaluationsliteratur handelt, soll der Unterschied in exakter Form dargestellt werden. Hierfür ist ein kurzer formaler Einschub notwendig (nach Blundell/ Costa Dias 2008).

Tabelle 18: Einteilung mikroökonomischer Evaluationsverfahren

Randomized Assignment	Selection on Observables	Selection on Unobservables
Soziales Experiment	Exaktes Matching	Differenz-von-Differenzen-Ansatz (Natürliches Experiment)
	Matching on the propensity score	Regression Discontinuity Design
	Hybrides Matching	Instrumentenvariablenansatz Kontrollfunktionsansatz

Eigene Darstellung

Wir spezifizieren hierzu folgende allgemeine Gleichung:

$$Y_i = \alpha + \beta_i D_i + u_i, \quad (14)$$

wobei Y die Ergebnisgröße und D den Teilnahmeindikator bezeichnen. Demnach bezeichnet β den Maßnahmeneffekt der Teilnahme. Das Subskript i deutet an, dass grundsätzlich jedes Unternehmen, welches eine Förderung erhält, einen anderen Maßnahmeneffekt aufweist (heterogene Maßnahmeneffekte). Der Erwartungswert von β , $E(\beta_i)$ bezeichnet jedoch einen anderen Evaluationsparameter als den bisher verwendeten MTT. Es handelt sich hierbei um den „Durchschnittlichen Maßnahmeneffekt“ (Average Treatment Effect, ATE) der definiert ist als

$$ATE = E\{Y_i(1) - Y_i(0)\}. \quad (15)$$

Der ATE bezeichnet dabei den Maßnahmeneffekt auf ein zufällig aus der Population ausgewähltes Unternehmen.

Weiterhin nehmen wir an, dass die Teilnahmeentscheidung eine Funktion der jeweils verfügbaren Information ist, die insbesondere auch die potenziellen Ergebnisse $Y(1)$ bzw. $Y(0)$ betrifft. Diese Information fassen wir zusammen in einen Vektor Z , der alle beobachtbaren teilnahmerelevanten Variablen enthält und einen Vektor ε , der alle unbeobachtbaren, teilnahmerelevanten Variablen enthält. Weiterhin definieren wir

$$D_i = 1 \text{ wenn } D_i^* \geq 0 \text{ und } D_i = 0 \text{ wenn } D_i^* < 0, \quad (16)$$

wobei D^* eine Funktion von Z und ε ist. D_i^* , ist dabei ein Ergebnis folgender Spezifikation:

$$D_i^* = 1[Z_i\gamma + \varepsilon_i \geq 0]. \quad (17)$$

Dabei bezeichnet γ einen Parametervektor. (17) ist ein sogenannte Indikatorfunktion, die den Wert 1 annimmt wenn die Bedingung in der Klammer erfüllt ist.

Wenn man in Gleichung (14) den Evaluationsparameter ATE (15) einsetzt, wird die Heterogenität der Maßnahmeneffekte Teil des Fehlerterms:

$$Y_i = \alpha + \beta^{ATE}D_i + [u_i + D_i(\beta_i - \beta^{ATE})] = \quad (18)$$

$$\alpha + \beta^{ATE}D_i + e_i. \quad (19)$$

Offensichtlich hat der Fehlerterm in (18), e , nunmehr eine komplexe Struktur. Ein Selection Bias liegt vor, wenn der Teilnahme-Indikator D mit dem Fehlerterm e korreliert ist, d.h. wenn

$$Cov(D, e) \neq 0. \quad (20)$$

Wenn Gleichung (20) erfüllt ist, hat dies weitreichend negative Konsequenzen für die Validität der Maßnahmenschätzung. Technisch gesprochen sind die Schätzer in diesem Fall verzerrt und inkonsistent, d.h. auch bei unendlich großer Stichprobe würden wir nicht den wahren Maßnahmeneffekt schätzen. Man bezeichnet D dann als endogene Variable. Diese Korrelation zwischen Maßnahmenindikator und Fehlerterm kann auf zwei Ursachen zurückgeführt werden, welche die Differenzierung in „Selection on Observables“ und in „Selection on Unobservables“ erklären. Um dies zu verdeutlichen, setzen wir in Gleichung (19) für den Teilnahmeindikator D die Indikatorfunktion (17) ein:

$$\alpha + \beta^{ATE}\{1[Z_i\gamma + \varepsilon_i \geq 0]\} + e_i. \quad (21)$$

Wie bereits erwähnt ist der Teilnahmeindikator das Ergebnis einer Entscheidung auf Basis von beobachtbaren und nicht beobachtbaren Variablen (siehe (17)). Demnach können wir eine Korrelation zwischen D und e nunmehr ausdifferenzieren in eine Korrelation zwischen Z bzw. ε und e .

Eine „Selection on observables“ liegt dann vor, wenn eine Korrelation zwischen Z und e die Ursache für die Korrelation zwischen D und e ist, d.h.

$$Cov(D, e) \neq 0 \text{ weil } Cov(Z, e) \neq 0. \quad (22)$$

Bei einer „Selection on unobservables“ ist D mit e korreliert weil eine Korrelation zwischen ε und e vorliegt, d.h.

$$Cov(D, e) \neq 0 \text{ weil } Cov(\varepsilon, e) \neq 0. \quad (23)$$

Tabelle 18 zeigt, dass jeweils nur bestimmte mikroökonomische Methoden in der Lage sind, diese unterschiedlichen Selektionsmechanismen zu berücksichtigen.

Der in der FTI-Evaluation und Arbeitsmarktevaluation dominante Matching-Ansatz kann nur für eine Selektion aufgrund beobachtbarer Merkmale kontrollieren, während z.B. Instrumentenvariablen-Schätzer oder Kontrollfunktionsansätze auch für Selektion aufgrund unbeobachteter Charakteristika kontrollieren. Es stellt sich nun die Frage, warum man nicht immer einen Schätzer anwendet, der auch für unbeobachtete Heterogenität, d.h. Selektion aufgrund unbeobachteter Merkmale, kontrolliert. Damit wäre man sozusagen immer auf der „sicheren Seite“. Die Antwort liegt darin, dass diese Schätzer wieder andere Nachteile aufweisen und typischerweise mehr theoretische Annahmen notwendig machen, deren Erfüllung wiederum fragwürdig sein kann.

Im Folgenden werden in knapper Form die Identifikationsannahmen von zwei idealtypischen Verfahren für diese beiden unterschiedlichen Typen von Selektionsprozessen dargestellt. Zunächst erfolgt eine Präsentation des Matching-Ansatzes, der eine Selektion aufgrund beobachtbarer Merkmale annimmt und danach eine Darstellung des Kontrollfunktionsansatzes.

Matching

Der Matching-Ansatz gehört aktuell zu den populärsten mikroökonomischen Methoden zur Evaluation politischer Maßnahmen in den Bereichen Arbeitsmarktpolitik und FTI-Politik. Eine wichtige Ursache für diese Popularität liegt zweifelsohne in der nachvollziehbaren Logik, die dieser Methode zugrunde liegt. Hier liegt ein großer Unterschied zu den weiter unten vorgestellten Kontrollfunktionsansätzen. Diese sind insbesondere bei formaler Darstellung nur schwer ohne ökonomische Vorkenntnisse verständlich. Ein weiterer Unterschied zwischen diesen beiden Ansätzen liegt darin, dass beim Matching-Verfahren eine Selektion aufgrund beobachtbarer Merkmale angenommen wird. Liegt diese Konstellation vor, so ist das Matching die idealtypische Methodenwahl. Das Matching-Verfahren gehört weiterhin zur Gruppe nichtparametrischer Verfahren und gilt deshalb als robust relativ zu hochparametrisierten Ansätzen wie z.B. Kontrollfunktionsansätzen. Trotzdem liegen auch beim Matching-Verfahren eine Reihe technisch anspruchsvoller Problemstellungen vor, die jedoch beim Konsumenten der Ergebnisse in der Regel nicht wahrgenommen werden.

Die Idee beim Matching-Ansatz ist zunächst einmal sehr einfach: Finde zu einem geförderten Unternehmen ein nicht gefördertes Unternehmen, welches als statistischer Zwilling für das geförderte Unternehmen fungiert. Es unterscheidet sich also lediglich beim Merkmal Förderung. Liegt diese ideale Datensituation vor, so kann die Zuteilung der Förderung auf das geförderte Unternehmen als zufällig aufgefasst werden, und ein direkter Vergleich der Erfolgsgröße zwischen diesen beiden Unternehmen ergibt den kausalen Maßnahmeneffekt. Damit wird auch die Ähnlichkeit dieses Ansatzes zum randomisierten Experiment deutlich. Die Identifikationsannahme ist ähnlich zu jener beim Querschnittsvergleich. Allerdings wird diese Annahme in zweierlei Hinsicht modifiziert. Erstens wird neben einer Übereinstimmung bei den Merkmalsausprägungen

der exogenen Variablen X auch eine Übereinstimmung bei vergangenen Werten der Erfolgsgröße Y angestrebt. Das bedeutet, dass etwa die Patentperformance der beiden statistischen Zwillinge in den Jahren vor der Förderung berücksichtigt wird.

Durch diese Berücksichtigung der „Vorgeschichte“ wird auch für jenen Anteil unbeobachtbarer Heterogenität kontrolliert, der sich in der Ausprägung dieser verzögerten Ergebnisgrößen niederschlägt. Zweitens erfolgt im Gegensatz zum Querschnittsvergleich keine „pauschale Kontrollgruppenbildung“ (Reinkowski 2005). Darunter versteht man die Tatsache, dass beim Querschnittsvergleich, der typischerweise als Regressionsmodell implementiert wird, ein Vergleich der geförderten Unternehmen mit allen Nichtteilnehmern erfolgt, wobei diese das gleiche Gewicht beim Vergleich erhalten. Dies ist beim Matching-Ansatz anders, da hier für jeden Teilnehmer ein passgenauer statistischer Zwilling identifiziert wird, der im einfachsten Fall mit dem Gewicht 1 in den Vergleich eingeht, während alle anderen Nichtteilnehmer in der Kontrollgruppe das Gewicht 0 erhalten. Es gibt dann auch noch eine Reihe weiterer Methoden, die eine Gewichtung als Funktion des Abstandes der Merkmalsausprägungen zwischen Teilnehmern und Nichtteilnehmern vornehmen. Während ersteres als Nearest-Neighbour-Matching bezeichnet wird, wäre etwa das Kernel-Matching ein Beispiel für letzteren Ansatz. Diese beiden Unterschiede zum einfachen Querschnittsvergleich erhöhen die Wahrscheinlichkeit, dass mittels Matching-Ansatz der wahre Maßnahmeneffekt geschätzt werden kann. Damit lautet die Identifikationsannahme:

$$E(Y_i(0)|X_i, Y_{i,t_0}, D_i = 1) = E(Y_i(0)|X_i, Y_{i,t_0}, D_i = 0). \quad (24)$$

Der Maßnahmeneffekt im Sinne des „mean effect of Treatment on the treated“ (MTT) ergibt sich dabei als gewichtete Differenz:

$$MTT_{Match} = \sum_{i=1}^{N_1} w(i)(Y_i(1) - Y_i(0)),$$

wobei $w(i)$ das jeweilige Gewicht beim Vergleich von Maßnahmenteilnehmer i mit der Kontrollgruppe angibt.

Das bisher beschriebene Verfahren stellt restriktive Anforderungen an das Datenmaterial und die Rechenkapazitäten. Eine Übereinstimmung bei allen exogenen Variablen und verzögerten Ergebnisgrößen wird tatsächlich nur selten zu erreichen sein. Aufgrund der genauen Übereinstimmung in den einzelnen Variablen wird diese Form des Matching auch als exaktes Matching bezeichnet. Um dieses Problem exakter Übereinstimmung zu umgehen, hat die methodische Forschung herausgestellt, dass auch eine alternative Form des Matching mit geringeren Anforderungen an das Datenmaterial ebenfalls zu validen Ergebnissen kommen kann. Es handelt sich hierbei um das sogenannte „Matching on the Propensity Score“. Dieses ist auch das in der Evaluation dominante Matchingverfahren. Es besteht aus zwei Schritten: In einem ersten Schritt werden für jedes Unternehmen aus Teilnehmer- und Kontrollgruppe Teilnahmewahrscheinlichkeiten berechnet. Diese Teilnahmewahrscheinlichkeit wird als „Propensity Score“ bezeichnet und gibt die bedingte Wahrscheinlichkeit zur Förderung, gegeben

den beobachtbaren Merkmalsvektor X , $P(X)=P(D=1|X)$, an. Diese Wahrscheinlichkeit wird mit nichtlinearen Regressionsmodellen, typischerweise einem Probitmodell geschätzt. Damit sind aber auch bei diesem grundsätzlich nichtparametrischen Verfahren Verteilungsannahmen notwendig.

In einem zweiten Schritt erfolgt das Matching zwischen diesen beiden Gruppen, d. h. die Zuordnung von nicht geförderten Unternehmen zu geförderten Unternehmen zwecks Bildung der Differenz der Ergebnisgrößen, auf Basis der Teilnahmewahrscheinlichkeit.

Es werden also Unternehmen mit ähnlicher Teilnahmewahrscheinlichkeit miteinander verglichen. Die Identifikationsannahme beim exakten Matching ändert sich beim „Matching on the Propensity Score“ entsprechend zu

$$E(Y_i(0)|P(X), D_i = 1) = E(Y_i(0)|P(X), D_i = 0). \quad (25)$$

Neben der Identifikationsannahme in (25) gibt es zwei weitere wesentliche Annahmen, die beim Matchingansatz erfüllt sein müssen, wobei die erste Annahme eine alternative Formulierung der Identifikationsannahme in Gleichung (25) darstellt.

- Conditional Independence Assumption oder Unconfoundedness Assumption. Dabei handelt es sich um die kritische Identifikationsannahme beim Matchingansatz. Sie besagt, dass, gegeben die beobachtbaren Merkmale X , die potenziellen Ergebnisse unabhängig von der Maßnahmenteilnahme sind. Anschaulich ausgedrückt: Für alle Unternehmen mit den gleichen Merkmalen X ist das potenzielle Ergebnis bei Förderung oder Nichtförderung identisch:

$$Y_i(1), Y_i(0) \perp D_i | X_i \text{ bzw. } Y_i(1), Y_i(0) \perp D_i | P(X_i). \quad (26)$$

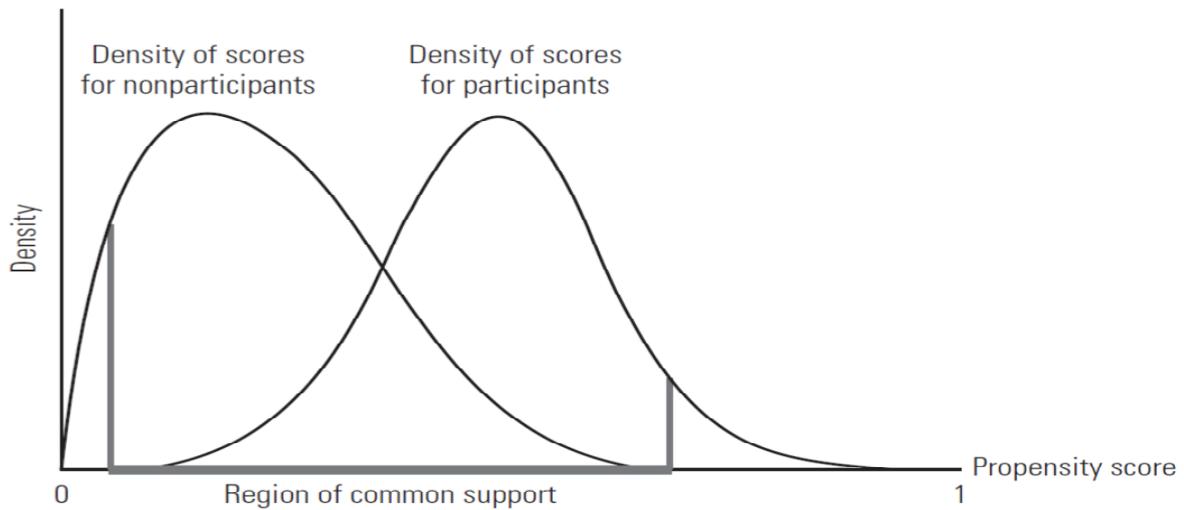
Diese Annahme trifft im Wesentlichen die Identifikationsannahme einer Selektion aufgrund beobachtbarer Merkmale X . Sie ist jedoch – wie die meisten kritischen Identifikationsannahmen – nicht testbar: „The acceptance of the unconfoundedness assumption cannot be directly tested, but the availability of ample information is important in order to define a vector of covariates X that makes the assumption more plausible.“ (Gonzalez/Pazo 2008:377)

- Comon Support Assumption oder Overlap Assumption. Diese Annahme verlangt dass für jedes geförderte Unternehmen mit Charakteristika X auch ein nicht gefördertes Unternehmen mit den gleichen Merkmalsausprägungen bzw. der gleiche Förderwahrscheinlichkeit existiert.

$$0 < P(D_i = 1|X_i) < 1 \quad (27)$$

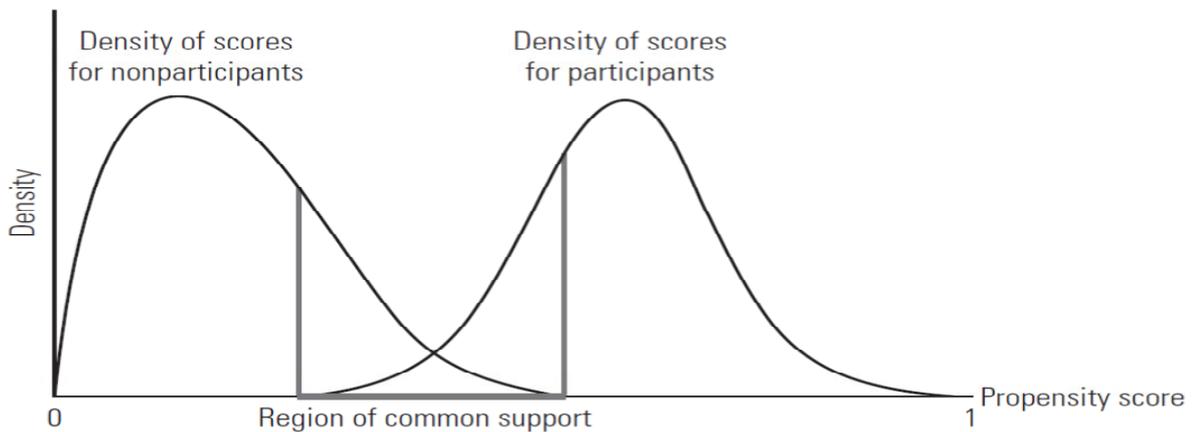
Dies ist eine notwendige Annahme, weil ansonsten keine passenden Matchingpartner gefunden werden können. Diese Annahme wäre verletzt, wenn z.B. alle Unternehmen einer Branche durch eine branchenspezifische Fördermaßnahme gefördert werden. Trifft dieser Fall zu, kann kein Matching-Ansatz angewandt werden. Für die Annahme (27) existieren Prüfverfahren. Abbildung 30 und Abbildung 31 illustrieren die Erfüllung bzw. Verletzung der Annahme. Dabei ist auf der Abszisse der Propensity Score aufgetragen, der Werte zwischen 0 und 1 annehmen kann (es handelt sich hierbei um eine Wahrscheinlichkeit). Auf der Ordinate wird der zugehörige Anteil bzw., weil es sich um stetige Zufallsvariablen handelt, die Dichte aufgetragen.

Abbildung 30: Common Support Assumption ist erfüllt



Khandker et al. 2010

Abbildung 31: Common Support Assumption ist möglicherweise verletzt



Khandker et al. 2010

Box 18 stellt zwei Beispiele vor, die Anwendungen des Matchingverfahrens in FTI-Evaluierungen zeigen. Nachdem es sich dabei um jeweils sehr unterschiedliche Förderinstrumente handelt wird deutlich, dass mikroökonomische Methoden einen sehr flexiblen Einsatz ermöglichen. Bei einer passenden Wahl der Ergebnisgröße können die unterschiedlichsten Instrumente mit ökonomischen Verfahren evaluiert werden. Kritisch sind letztlich vor allem die Anzahl der Programmteilnehmer sowie die Verfügbarkeit einer Kontrollgruppe.

Box 18: Matchingansätze in der FTI-Evaluation

Beispiel 1: FlexCIM-Evaluierung in Österreich (Polt/ Pointner 2005).

Das Programm FlexCIM wurde 1991 eingeführt um die Wettbewerbsfähigkeit der österreichischen Unternehmen durch den Einsatz von flexiblen computergestützten Produktionsprozessen zu erhöhen. Gefördert wurde die Einführung und damit die Diffusion einer Produktionstechnologie. Das Programm lief 5 Jahre bis 1996. Bei der Evaluation wurde zum ersten Mal in Österreich ein (Propensity-Score-)Matching-Ansatz angewandt. In einem ersten Schritt wurden die Teilnahmewahrscheinlichkeit der 84 Teilnehmer und 217 Kontrollunternehmen geschätzt. Dazu wurde die folgende Funktion geschätzt:

$$P(X) = f \left\{ \begin{array}{l} \text{Region des Unternehmens, Betriebsalter, Anzahl der Beschäftigten,} \\ \text{Zwegbetrieb in ausländischem Mehrheitsbesitz, FE – Quote,} \\ \text{Fertigungsverfahren, Wettbewerbsmodus auf den Absatzmärkten} \end{array} \right\}$$

Anschließend erfolgte das Matching auf Basis der ermittelten Teilnahmewahrscheinlichkeiten sowie unter Berücksichtigung von vier Branchen, d.h. es wurde nur jene Unternehmen miteinander verglichen, die der gleichen Branche angehören und einen ähnlich hohen Propensity Score $p(X)$ aufweisen. Daran anschließend wurden die Mittelwerte verschiedener Ergebnisgrößen von Teilnehmer- und Kontrollgruppe miteinander verglichen und ein Test auf signifikanten Mittelwertunterschied (t-Test) gerechnet. Tabelle 19 zeigt einen Auszug aus den Ergebnissen am Beispiel des Vergleichs der Ergebnisgröße „Anzahl CIM-Elemente“. Dabei zeigt sich in der letzten Zeile, dass die geförderten Unternehmen signifikant mehr FlexCIM-Elemente zwischen 1991-1999 einführten als die nicht geförderten Unternehmen.

Tabelle 19: Ergebnisse der FlexCIM-Evaluierung

Ergebnisgröße	Programtteilnehmer	Nicht-Teilnehmer	Differenz (%-Punkte)
Anzahl CIM-Elemente 1999	6,7	6,2	0,5
Anzahl CIM-Elemente 1991	2,2	2,5	-0,3
Zunahme CIM-Elemente 1991-1999	4,5	3,7	0,7*

* Signifikant am 0,05-Niveau

Beispiel 2: CTI-Evaluierung in der Schweiz (Arvanitis 2010).

Arvanitis et al. (2010) untersuchen die kausalen Fördereffekte durch die schweizerische Förderagentur CTI am Beispiel von 199 zwischen 2000-2002 geförderten Unternehmen. Dabei verwenden sie vier unterschiedliche Matchingverfahren (Nearest Neighbour Mmatching, Calliper Matching, Kernel Matching, Local Linear Regression Matching), um die Robustheit der Ergebnisse zu erhöhen. Dieses Matching-Verfahren wenden sie auf sechs unterschiedliche Ergebnisgrößen an, um möglichst viele potenzielle Maßnahmeneffekte zu erfassen. Dazu schätzten die AutorInnen in einem ersten Schritt den Propensity Score anhand folgender Funktion:

$$P(X) = f \left\{ \begin{array}{l} \text{Unternehmensgröße, kontinuierliche FE – Aktivitäten,} \\ \text{Exportaktivitäten, Gründungsjahr, Branche, Region, Sprache des Fragebogens} \end{array} \right\}$$

Tabelle 20 stellt die Ergebnisse für das Nearest Neighbour Matching dar. Zunächst einmal zeigt sich in der Differenz zwischen den Werten von Spalte 1 und 2 die Auswahl von überdurchschnittlich innovativen Unternehmen als Vergleichsgruppe für die geförderten Unternehmen. Die letzte Spalte zeigt schließlich, dass die Förderung auf 5 der 6 Ergebnisgrößen einen positiven und signifikanten Effekt hatte.

Tabelle 20: Ergebnisse des Matchingverfahrens der CTI-Evaluierung (Nearest Neighbour Matching)

Ergebnisgröße	Gesamte Kontrollgruppe (1)	Kontrollgruppe nach Matching (2)	Geförderte Unternehmen (3)	Differenz (3)-(2)
Bedeutung der eingeführten Innovation aus technologischer Perspektive	3,34	3,44	3,75	0,31*
Bedeutung der eingeführten Innovation aus ökonomischer Perspektive	3,36	3,60	3,65	0,005
Prozentuale Reduktion der durchschnittlichen Variablen durch Prozessinnovation	4,98	3,59	8,61	5,02*
Umsatzanteil von signifikant verbesserten oder modifizierten Produkten	33,73	36,60	48,36	11,76*
Umsatzanteil von Produkten die neu für das Unternehmen oder den Markt sind	15,73	17,24	27,46	10,22*
Umsatzanteil von Produkten die neu für die Welt sind	4,44	3,01	15,58	12,57*

* Signifikant am 0,05-Niveau

Kontrollfunktionsansatz

Im Gegensatz zum Matching-Ansatz liegt der Vorteil des Kontrollfunktionsansatz darin, dass auch für eine Selektion aufgrund unbeobachteter Merkmale kontrolliert werden kann. Nachdem eine formale Darstellung dieses Ansatzes zu komplex für den Zweck dieser Studie ist, soll im Folgenden lediglich eine intuitive Präsentation davon erfolgen, wie eine Kontrolle für eine „Selection on Unobservables“ möglich ist.¹⁵

Der Kern des Kontrollfunktionsansatzes wird durch eine Modellierung des Selektionsprozesses gebildet. Dabei wird unterstellt, dass sich die Teilnehmer als Homo Oeconomici verhalten. Ein Unternehmen nimmt nur dann an der Förderung teil, wenn die Differenz zwischen Kosten und Nutzen positiv ist, wobei der Nutzen etwa in der Veränderung des erwarteten Gewinns besteht.

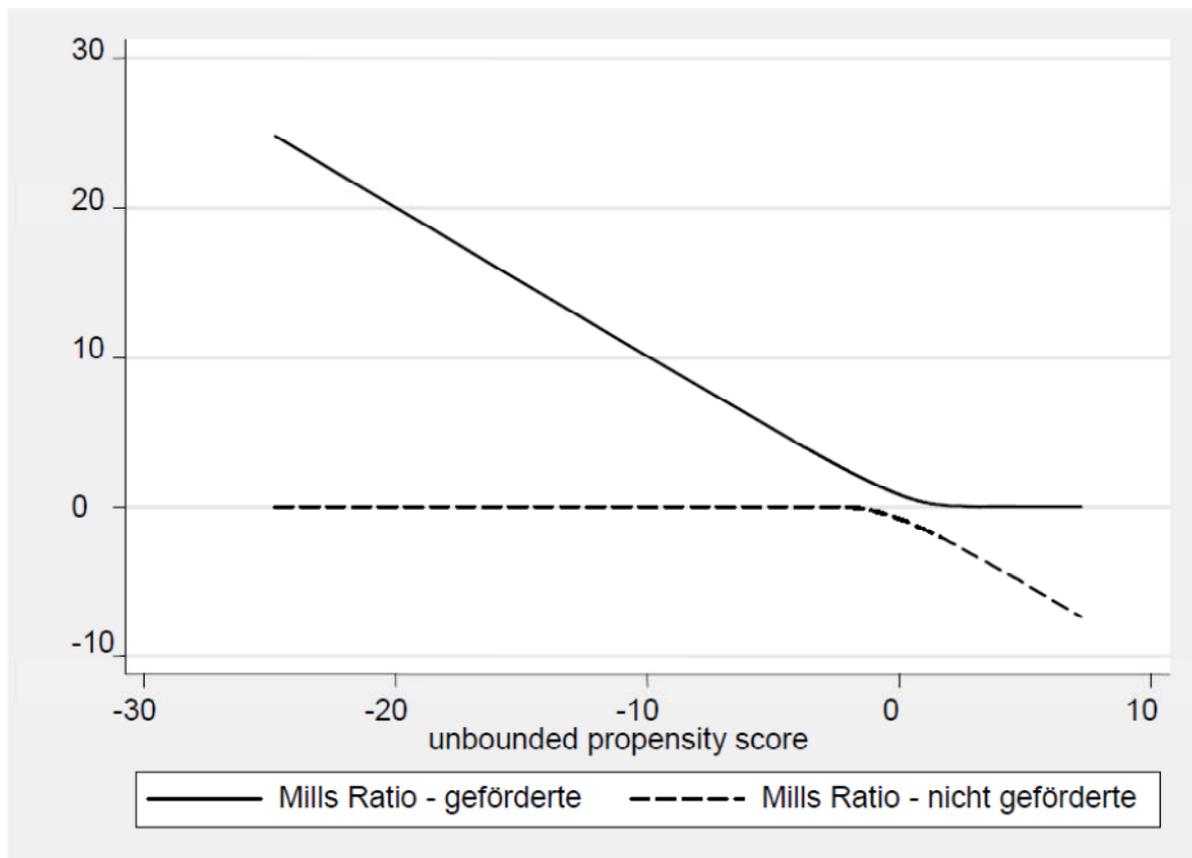
¹⁵ Für das Folgende vgl. Bauer et al. (2009). Der ökonomisch interessierte Leser sei auf die beiden Lehrbücher von Greene (2011) sowie Cameron und Trivedi (2005) verwiesen.

Gemäß der bisherigen Notation muss folgende Ungleichung erfüllt sein:

$$[Y_i(1) - Y_i(0)] - C > 0, \quad (28)$$

wobei C die durch die Teilnahme entstehenden Kosten bezeichnet. Wir können diese Gleichung für die Teilnehmer nicht ausrechnen, weil wir keine Informationen über die kontrafaktische Situation $Y(0)$ und über die Teilnahmekosten haben. Die Teilnahmeentscheidung ist wiederum eine Funktion von beobachtbaren Variablen X und unbeobachtbaren Variablen ε (siehe oben). Angenommen Managementqualität ist eine unbeobachtbare Variable, welche sowohl die Teilnahmewahrscheinlichkeit als auch die Ergebnisgröße positiv beeinflusst. Weiterhin kann davon ausgegangen werden, dass diese Variable vor allem auch einen Einfluss auf die beiden unbekanntenen Größen $Y(0)$ und C in Gleichung (28) hat. Demnach wäre eine mögliche Hypothese etwa ein positiver Zusammenhang von Managementqualität und der kontrafaktischen Situation bei Nichtteilnahme, $Y(0)$, sowie ein negativer Zusammenhang mit den Teilnahmekosten, C. Damit ermöglicht Entscheidungsgleichung (28) die Gewinnung von Informationen über eigentlich unbeobachtbare Größen. Wenn nun beispielsweise ein Unternehmen eine Förderung erhält, dies aber aufgrund der beobachtbaren Merkmalsausprägungen der Variablen X eher unwahrscheinlich ist, so ist z.B. anzunehmen, dass eine höhere Managementqualität des betreffenden Unternehmens vorliegt, die für die Erfüllung der Ungleichung (28) sorgt. Durch diese Information kann ein Korrekturterm konstruiert werden und bei der Schätzung der Maßnahmeneffekte berücksichtigt werden. Dieser Korrekturterm wird meist als „Mills Ratio“ bezeichnet. Abbildung 32 zeigt die Funktion dieses Korrekturterms. Auf der Abszisse ist die Teilnahmewahrscheinlichkeit aufgetragen und auf der Ordinate der Wert des Korrekturterms. Demnach erfolgt bei geförderten Unternehmen, deren beobachtbaren Merkmale eigentlich gegen eine Förderung sprechen würden, eine Korrektur „nach oben“, weil offenbar unbeobachtete Charakteristika eine Rolle bei der Selektion spielten.

Abbildung 32: Korrektur beim Kontrollfunktionsansatz



Czarnitzki et al. 2003

Die konkrete Implementierung geschieht in der Regel in einem zweistufigen Verfahren, wobei in einem ersten Schritt eine Selektionsgleichung über ein Probitmodell geschätzt wird. Daraus wird ein Korrekturterm errechnet und in einem zweiten Schritt in einem linearen Regressionsmodell, mit der Ergebnisgröße als abhängiger Variable, berücksichtigt. Dabei werden regelhaft zwei kritische Annahmen getroffen: Erstens wird angenommen, dass die Störterme der beiden Regressionsmodelle gemeinsam normalverteilt sind (Verteilungsannahme). Zweitens benötigt man in der Regel eine sogenannte Ausschlussrestriktion, d.h. es sollte zumindest eine exogene Variable der Selektionsgleichung nicht in der zweiten Stufe der linearen Regression aufgenommen werden. Freilich ist es alles andere als leicht, eine Variable zu finden, die zwar die Teilnahme, aber nicht die Ergebnisgröße beeinflusst. Diese beiden Annahmen führen letztlich dazu, dass Matchingansätze als robuster wahrgenommen werden im Vergleich zu Kontrollfunktionsansätzen. Abbildung 33 fasst die Vor- und Nachteile von Matching- und Kontrollfunktionsansätzen zusammen.

Abbildung 33: Vor- und Nachteile von Matching- und Kontrollfunktionsansätzen im Vergleich

	Vorteile	Nachteile
Selektionsmodelle (Kontrollfunktions- methoden)	<ul style="list-style-type: none"> - Die Selektion in Programme kann auch von unbeobachteten Charakteristika abhängen. - Es wird kein "Common Support" benötigt, d.h. die Verteilung der Teilnahmewahrscheinlichkeiten der Teilnehmer und Nichtteilnehmer muss sich nicht komplett überlappen. 	<ul style="list-style-type: none"> - Bestimmte funktionale Form der Untersuchungsgleichung muss angenommen werden (parametrisch); in der Regel linear: $y_i = X_i\beta + d_i\theta + u_i$ - Es wird eine "Ausschlussrestriktion" benötigt, d.h. mindestens ein Merkmal, dass die Programmteilnahme (wesentlich) beeinflusst, nicht aber die Zielvariable.
Matching Methoden	<ul style="list-style-type: none"> - Funktionale Form der Untersuchungsgleichung ist vollkommen flexibel (nichtparametrisch): $y_i = g(X_i) + d_i\theta + u_i$ - Es wird keine "Ausschlussrestriktion" benötigt. 	<ul style="list-style-type: none"> - Man nimmt an, dass die Selektion in Programme nur aufgrund von beobachtbaren Charakteristika erfolgt. - Es wird "Common Support" benötigt, d.h. für jeden Teilnehmer müssen Nichtteilnehmer mit ähnlicher Teilnahmewahrscheinlichkeit vorhanden sein.

Czarnitzki et al. (2003)

Die Anzahl der Anwendungen von Kontrollfunktionsansätzen ist wesentlich geringer als die Anzahl der Anwendungen von Matching-Ansätzen bei FTI-Evaluierungen. Ein Grund hierfür liegt im relativ komplexen Teilnahmeprozess bei F&E-Förderungen. Clarysse und Knocaert (2009) diskutieren dieses Problem und kommen zum Schluss, dass die Faktoren, welche die Unternehmensentscheidung zum Förderantrag beeinflussen, andere sind, als jene Faktoren welche die Förderentscheidung durch die Förderagentur bestimmten. Offensichtlich berücksichtigt Entscheidungsgleichung (28) noch nicht alle relevanten Aspekte. Ein Anwendungsbeispiel zur Evaluation von Kontrollfunktionsansätzen zur Evaluation von innovationspolitischen Maßnahmen findet sich in Hussinger (2008).

5.2.5 Grenzen ökonometrischer Modellierung

Die Nachteile ökonometrischer Evaluation liegen zum Teil gerade dort, wo die qualitativen Methoden ihre Stärken haben. Dies betrifft etwa die Frage nach den konkreten Mechanismen einer Förderwirkung in einem Unternehmen, die durch qualitative Fallstudien analysiert werden könnten. Spezifische Nachteile betreffen vor allem folgende Aspekte:

- Es werden im Rahmen der mikroökonometrischen Evaluation allgemeine Gleichgewichtseffekte ausgeschlossen (Stable-Unit-Treatment-Value-Assumption). Diese können jedoch gerade bei einer hohen Förderbreite und Förderintensität relevante Ausmaße annehmen. Ebenfalls unberücksichtigt bleiben auch Spillovereffekte, was zu paradoxen Ergebnissen führen kann: „To conclude, we face the paradoxical situation that if an evaluation study finds little difference between the supported firms and the non-supported firms it could be either be because the R&D program was unsuccessful and generated little innovation or because the R&D program was highly successful in generating new innovations which created large positive spillovers to the non-supported forms.“ (Klette et al. 2000:482).

- Damit verbunden ist auch die Tatsache, dass nur Aussagen zur Effektivität, aber nicht zur Effizienz eines Förderprogramms möglich sind. Dies ist eine Folge mangelnder Erfassung von Verdrängungs- und Substitutionseffekten. Eine Erfassung dieser ist jedoch mit einer signifikanten Zunahme an methodischer Komplexität verbunden. Weiterhin sei bedacht, dass Effektivität eine notwendige Voraussetzung für Effizienz ist. Eine ineffektive Maßnahme kann – gegeben ein bestimmtes Ziel - niemals effizient sein.
- Eine Schwäche besteht freilich darin, dass gerade die entscheidenden Identifikationsannahmen nicht getestet werden können. Genau diese sind aber dafür ausschlaggebend, ob der Maßnahmenschätzer valide Ergebnisse liefert oder nicht.
- Eine ökonometrische Evaluation bietet vor allem summativ-evaluatorische und weniger formativ-evaluatorische Information. Allerdings sollte auch hier gefragt werden, welchen Wert detaillierte Kenntnis etwa über das Programmmanagement hat, wenn nicht einmal klar ist, ob das Programm seine Ziele erreicht oder nicht.
- Die Anforderungen an die Daten sind hoch. Beobachtungen über eine Kontrollgruppe sind notwendig und die Variablen müssen eine Modellierung des Selektionsprozesses oder eine Schätzung der Teilnahmewahrscheinlichkeit ermöglichen.
- Ökonometrische Verfahren haben in der Vermittlung gegenüber qualitativen Verfahren den Nachteil, dass sie tendenziell eine höhere Kompetenz auf Seiten des Konsumenten der Ergebnisse verlangen. Dabei können zwei unerwünschte Effekte auftreten: Einerseits kann es passieren, dass einzelne Werte eine überproportionale und unkritisch hinterfragte Bedeutung erlangen, andererseits können die Ergebnisse einfach abgelehnt werden, weil ein grundsätzliches Nicht-verstehen der Methode vorliegt.

Trotz dieser Nachteile mikroökonomischer Verfahren sollte deren fundamentaler Vorteil nicht übersehen werden. Dieser liegt darin, dass die letztlich entscheidende Frage nach der Kausalität eines Förderprogramms explizit gestellt wird. Welchen Nutzen kann eine Evaluation ohne Antwort auf diese Frage haben? Nicht zuletzt könnte damit etwa die folgende Feststellung von Tichy (2009:264) einem methodisch rigorosen Falsifikationsversuch unterzogen werden: „Es kann mit großer Sicherheit geschlossen werden, dass ein großer Teil der österreichischen Innovationsförderung von Projekten in Anspruch genommen wird, die ohne Förderung von den Firmen selbst finanziert worden wären und eher Weiterentwicklungen bestehender Produkte sind. Nur ein bescheidener Teil finanziert wirklich innovative Projekte, die die Wettbewerbsfähigkeit in ein oder zwei Jahrzehnten stützen werden.“

5.3. BESTIMMUNGSGRÜNDE DER METHODENWAHL

Welche Methoden sollen zur Evaluation für welches Förderprogramm verwendet werden? Diese Frage wurde bereits in mehreren Studien gestellt (e.g. Rhomberg et al. 2006). Dazu ist zunächst festzustellen, dass sich eine gute Evaluation durch eine an den jeweils spezifischen evaluatorischen Kontext angepasste Methodenselektion auszeichnet. Die Wahl der Methoden ist die vornehmliche Aufgabe der EvaluatorInnen und sollte etwa nicht in den Terms of References festgelegt werden.

Es wird dementsprechend in der einschlägigen Literatur die Meinung vertreten, dass es kein „Kochrezept“ gibt, welches für alle Fälle Gültigkeit beanspruchen könnte: „There is inherently no method of choice for conducting program evaluations.“ (Heckman et al. 1999:1868). Ähnlich argumentiert auch Reinkowski (2005:188) unter Bezug auf ökonomische Verfahren: „Allerdings kann keine allgemeingültige Handlungsempfehlung für die Verfahrensauswahl gegeben werden, da die Grundannahmen der vorgestellten Verfahren i.d.R. nicht testbar sind.“ Insofern ist die Festlegung von Methoden auf dem „Reisbrett“ ein durchwegs fragwürdiges Unterfangen, weil gute Evaluation kontextsensibel sein sollte.

Trotzdem gibt es einige Faustregeln, die jedoch unter dem bereits erwähnten Vorbehalt zu betrachten sind. Bei all dem ist zu betonen, dass letztlich vor allem die Datenlage sowie die Budgetbeschränkung darüber entscheiden, welche Methode zum Einsatz kommen kann oder nicht. Einige dieser „Faustregeln“ werden im Folgenden erörtert.

Eine Hilfestellung bei der Methodenselektion bietet etwa Tabelle 21 an. Dabei sind in den Zeilen unterschiedliche F&E-Instrumente eingetragen und in den Spalten die Methoden. Die Anzahl der Punkte zeigt den Grad an Adäquatheit an. Auffallend ist hier etwa die Tatsache, dass Kosten-Nutzen-Analyse als für viele Instrumente passend beurteilt wird aber de facto kaum eingesetzt wird – jedenfalls nicht in Europa.

Tabelle 21: Innovationspolitische Instrumente und adäquate Innovationsmethoden

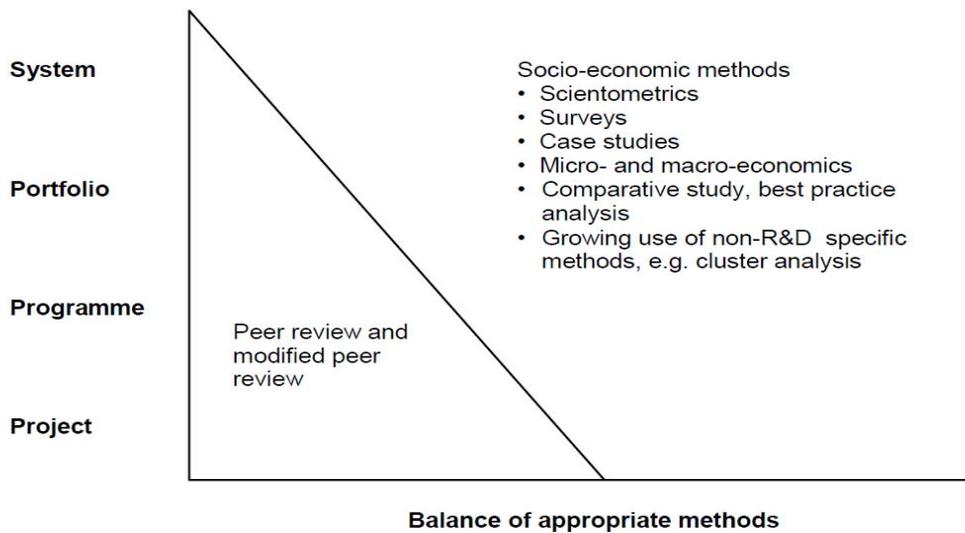
	Innovation Surveys	Econometric Models	Control Group Approaches	Cost Benefit Analysis	Expert Panels/ Peer Review	Field / Case Studies	Network Analysis	Foresight/ Technology Assessment	Benchmarking
Financing R&D	●●●	●●●	●●●	●		●●●		●	●
Provision of R&D infrastructure		●●		●●●	●●●	●●●	●●●	●●	●●●
Technology transfer and innovation diffusion	●●●	●●●	●●●	●●	●	●●	●●●	●●●	●●●
Legal frameworks (IPRs, standards and regulation)	●	●	●	●●●		●●●		●●	●●●
Integrated projects			●	●●●	●●●	●●●	●●●	●●	●●
Networks of excellence					●●●	●●	●●●	●●	●●

Legend: ●●● Highly suitable ●● Suitable ● Less suitable

RTD-Evaluation Toolbx 2002

Arnold (2005) stellt in seinem Aufsatz zur Methodik von Systemevaluierungen ein Schema zur Methodenselektion vor, welches die Methodenwahl in Abhängigkeit von der zu analysierenden Ebene im Innovationssystem präsentiert. Abbildung 34 zeigt dieses Schema, wobei auf der Ordinate eine Einteilung in vier Ebenen vom konkreten F&E-Projekt bis zum nationalen Innovationssystem getroffen wird. Demnach ist Peer Review als Evaluationsmethode immer schlechter geeignet, je höher die jeweils zu evaluierende Ebene angesiedelt ist. Gleichzeitig gewinnen ökonomische Methoden, Fallstudien etc. an Bedeutung.

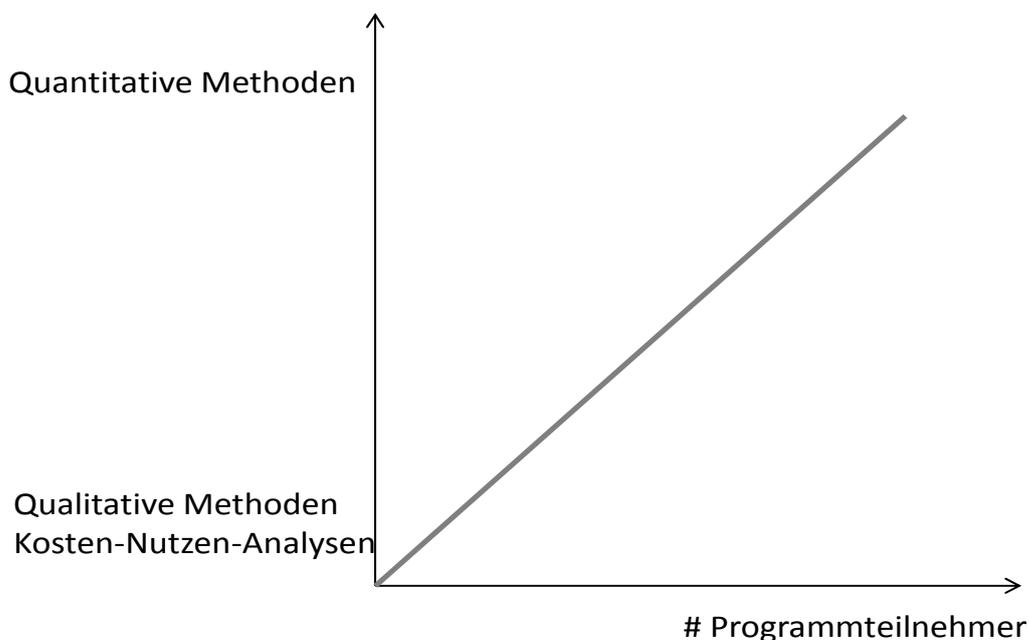
Abbildung 34: Evaluationsmethoden und Analyseebene



Arnold 2005

Betrachtet man die Methodenwahl als Funktion der Anzahl der Programmteilnehmer, so zeigt Abbildung 35 in schematischer Form einen positiven Zusammenhang zwischen der Anzahl der Programmteilnehmer und dem Grad an Quantifizierung. Geringe Fallzahlen sollten besser mit qualitativen oder einfachen deskriptiven Verfahren analysiert werden. Generell erscheint eine Fallzahl von 50-100 als eine notwendige Voraussetzung für mikroökonomische Anwendungen. Allerdings ist das keineswegs als wissenschaftliche Regel aufzufassen. Bei niedrigen Fallzahlen können auch Kosten-Nutzen-Analysen durchgeführt werden. Diese sind für die Bewertung einiger weniger alternativer Projekte konzipiert.

Abbildung 35: Der Zusammenhang von Methode und Anzahl der Programmteilnehmer



Eigene Darstellung

Als weitere Faustregel der Methodenwahl kann auch der Zusammenhang zwischen der zu messenden Additionalität und der Methodenwahl gelten. Demnach sind die Dimensionen der Input- und Outputadditionalität sehr gut mit ökonomischen Verfahren zu schätzen, während Verhaltensadditionalität eher mit qualitativen bzw. deskriptiven Methoden evaluiert werden sollte (Falk 2006). Allerdings ist auch hier keineswegs eine einheitliche Meinung in der Literatur zu finden und es gibt auch ökonomische Studien zur Verhaltensadditionalität (e.g. Marino/ Parrotta 2010).

In der Literatur wird auch vereinzelt darauf hingewiesen, dass ökonomische Methoden eher für ex-post Evaluierungen geeignet seien, während Interimsevaluation eher nach qualitativen Methoden verlangen würde. Dieses Argument ist in dem Sinne zutreffend, als bei Interimsevaluierungen auch das Lernen im Sinne einer Verbesserung des Programms von hoher Bedeutung ist und die dafür notwendige formative Information eher nicht mittels ökonomischen Verfahren ermittelt werden kann. Jedoch stellt sich die Frage, ob es Sinn macht, ein Programm in Details zu verbessern, ohne die grundsätzliche Frage der Zielerreichung zu stellen. Zusammengenommen kann festgestellt werden, dass eine Reihe von Hinweisen existieren, unter welchen Umständen welche Evaluierungsmethode eingesetzt werden sollte. Diese sollten jedoch keineswegs in schematischer Art und Weise angewandt werden. Vielmehr ist zweierlei zu beachten: (1) Gute Evaluation ist immer durch einen Methodenmix gekennzeichnet. (2) Gute Evaluation ist kontextsensibel und passt ihren Instrumenteneinsatz stets an die spezifische evaluatorische Aufgabe an. Wenn es harte Faktoren der Methodenwahl gibt, so betreffen diese die Datenverfügbarkeit und vor allem die Budgetbeschränkung. In diesem Sinne besteht ein positiver Zusammenhang zwischen der Zahlungsbereitschaft der öffentlichen Hand für Evaluierungen und dem Einsatz kreativer und innovativer Methoden. Diese Zahlungsbereitschaft betrifft dabei wohlgerne nicht die Gesamtausgaben für die Summe aller Evaluierungen sondern die Ausgaben pro Evaluation. So wird etwa eine Aufteilung der gleichen Summe auf 10 kleine Evaluierungen gänzlich andere Evaluationsmethoden hervorbringen als eine Aufteilung dieser Summe auf drei große Evaluierungen.

6 Sozio-ökonomische Wirkungsdimensionen von FTI-Maßnahmen

“As the notion of innovation policy becomes broader and more inclusive, measuring innovation across different policy domains is a huge challenge that calls for reconsidering the framework for measuring innovation.”

OECD Innovation Strategy (2010:205)

6.1. DEFINITION SOZIO-ÖKONOMISCHER WIRKUNGSDIMENSIONEN

Unter sozio-ökonomischen Wirkungsdimensionen werden die unterschiedlichsten gesellschaftspolitischen und ökonomischen Ziele subsumiert, welche durch FTI-Fördermaßnahmen angestrebt werden.

Wirkungstypen erfassen die klar umgrenzten, messbaren Wirkungen von FTI-Maßnahmen und können in zwei verschiedene Arten erhoben werden (vgl. Falk 2006):

- Ergebnisorientierte Wirkungstypen: kennzeichnen sich durch ihre eindeutige Erfassbarkeit in Zahlenwerten (Schaffung neuer Arbeitsplätze, Anzahl an Publikationen etc.)
- Verhaltensorientierte Wirkungstypen: zielen auf sozio-ökonomische Veränderungsprozesse ab (Veränderungen im Innovationverhalten, Imageveränderung etc.)

Diese Unterscheidung der Wirkungstypen existiert vorwiegend „theoretisch“ bei der Definition der Ziele z.B. in den Programmdokumenten; in der Evaluationspraxis findet zumeist eine Einengung auf wenige zentrale Indikatoren statt, die sowohl zur Operationalisierung ergebnisorientierter als auch verhaltensorientierter Wirkungstypen eingesetzt werden. Am Beispiel der Anzahl der Publikationen lässt sich dies sehr gut darstellen: Dieser Indikator zeigt zum einen das unmittelbare „Ergebnis“ eines Projekts, zum anderen aber verhaltensorientierte Wirkungen an (z.B. die Anzahl an Publikationen steigert die wissenschaftliche Reputation, etc.)

Godin und Doré (2006) zeigen mit ihrer Auflistung der „eleven dimensions of the impact of science“ die große Bandbreite sozio-ökonomischer Wirkungsdimensionen auf (Tabelle 22).

Tabelle 22: Die Bandbreite sozio-ökonomischer Wirkungsdimensionen

Science impacts: Knowledge, Research activities, Training	Organization impacts: Planning, Work organization, Administration, Human resources
Technology impacts: Products, Processes, Services, Know-how	Health impacts: Public health, Health systems
Economy impacts: Production, Financing, Investments, Commercialisation, Budget	Environment impacts: Management of natural resources and the environment, Climate and meteorology
Culture impacts: Knowledge, Know-how, Attitudes, Values	Symbolic impacts: Legitimacy/credibility/visibility
Society impacts: Welfare, Discourses and actions of groups	Training impacts: Curricula, Pedagogical Tools, Qualifications, Graduates, Insertion into the job market, Fitness of training/work, career, use of acquired knowledge
Policy impacts: Policymakers, Citizens, Public programs, National security	

Godin und Doré, 2006

Eine Zuordnung von Indikatoren zu den einzelnen Wirkungsdimensionen wurde von Godin und Doré (2006) beispielhaft skizziert und findet sich im Anhang A.

6.2. WELCHE SOZIO-ÖKONOMISCHEN WIRKUNGSDIMENSIONEN WERDEN BEI DER EVALUIERUNG VON FTI-MASSNAHMEN GENERELL ERHOBBEN?

Die große Bandbreite an anspruchsvollen FTI-relevanten Programmzielen steht einer deutlichen Einengung auf ökonomische, technologische und wissenschaftliche Wirkungen als einzig relevante Evaluierungskriterien von FTI-Programmen gegenüber. Sozio-ökonomische Wirkungsdimensionen, die auf verkehrs-, nachhaltigkeits-, energie-, und sicherheitspolitische sowie soziale und kulturelle Effekte abzielen, werden eher selten erhoben.

Box 19: Gründe für die Dominanz der ökonomischen Wirkungsdimensionen

- Godin und Doré nennen drei Gründe für die Dominanz der ökonomischen Dimensionen:
1. most measurement of science and research has been undertaken in an economic context
 2. most of the outputs and impacts of science are intangible, diffuse and often occur with important lags
 3. although also difficult to measure, the economic dimension of science and technology remains the least difficult

(Godin/Doré, 2006: 4)

In den letzten Jahren wurde verstärkt versucht, nicht-ökonomische Wirkungsdimensionen bei der Evaluierung FTI-politischer Maßnahmen zu berücksichtigen.

Gerade mit dem Aufkommen systemisch-struktureller FTI-Interventionen fließen neue Wirkungsdimensionen in den Evaluationsprozess ein, wie z.B. der Beitrag von FTI-Interventionen für nachhaltige Wissenschafts-Industrie-Beziehungen oder die Förderung der F&E Internationalisierung österreichischer KMUs. Hinzu kommen neue missionsgetriebene FTI-Politikmaßnahmen, die von Anfang an sozio-ökonomische Veränderungsprozesse als Wirkungsziele für sich definiert haben, wie z.B. die Steigerung der Lebensqualität älterer Menschen (benefit), der Abbau von Unsicherheiten in Bezug auf Umwelt- und Gesundheitsrisiken von Nanotechnologie (Nano) oder die CO₂ Reduktion durch Förderung erneuerbarer Energietechnologien (Neue Energien 2020).

Auch das Konzept der „sozialen Innovation“ gewinnt vermehrt an Bedeutung. Soziale Innovationen sind Lösungsstrategien, die in so genannten „communities of practice“ entwickelt werden – der Markt, der „user innovations“ und „producer innovations“ festhält, ist dabei aber nicht präsent; soziale Innovationen werden daher, trotz ihrer weitreichenden sozio-ökonomischen Bedeutung, durch reguläre ökonomische Standard-Indikatoren nicht erhoben (vgl. Gault 2011).

Obwohl ökonomische Indikatoren insbesondere für FTI-politische Maßnahmen interessant und nicht verzichtbar sind, repräsentieren sie nur die „eine Seite der Medaille“. Der Einbezug sozialer, gesellschaftspolitischer und kultureller Aspekte ermöglicht eine umfassende Analyse von Produkt- und Prozessinnovationen mit der Folge, dass:

- zum einen die Akzeptanz für die öffentliche Förderung erhöht werden kann, und
- zum anderen eine fundierte und umfassende Grundlage zur Legitimation/Abschaffung von Programmen zur Verfügung steht, da die Realisierung von FTI-Programmen oftmals in höherem Maße von Zielen anderer Politikfelder abhängt, als den im engeren Sinne FTI-politischen.

6.3. HERAUSFORDERUNGEN DURCH DEN EINBEZUG NICHT-ÖKONOMISCHER WIRKUNGSDIMENSIONEN BEI DER EVALUIERUNG VON FTI-MASSNAHMEN

Vier Herausforderungen für Wirkungsanalysen werden in der Folge beispielhaft beschrieben:

Zum ersten stellt sich für Wirkungsanalysen das zentrale Problem der zur Verfügung stehenden Indikatoren. In Hinblick auf Wirkungstypen, die über ökonomische Effekte hinausgehen, gibt es noch keine allgemein erprobten Indikatoren; bei sozialen und gesellschaftspolitischen Wirkungen handelt es sich zumeist um schwer messbare und nur langfristig beobachtbare Effekte.

Zum zweiten können nicht-ökonomische Effekte äußerst komplexe Auswirkungen auf die Gesellschaft haben; hier wird es schwierig, Wirkungsverkettungen zu identifizieren, messbar zu machen und zu analysieren.

Zum dritten besteht die Herausforderung, „Wertungen“ für nicht-ökonomische Wirkungen zu finden. Versuche gingen bisher in die Richtung, nicht-ökonomischen Effekten einen monetären

Wert zuzuordnen (z.B. ökonomische Ersparnisse, die durch Verkehrssicherheitsforschung produziert werden – vgl. Vinnova 2007).

Zum vierten zeigt sich, dass die wenigen Studien, die die Wirkung nicht-ökonomischer Zieldimensionen evaluieren, sich auf indirekte Effekte fokussieren. Dies ist eine Folge der Erhebungsmethodenwahl. Aufgrund der weit gefassten gesellschaftspolitischen Ziele werden vorwiegend offene und qualitativ angelegte Ansätze (z.B. ExpertInnengruppen, historische Fallstudien) verwendet, um den Wirkungen Rechnung zu tragen. Dies ist auch aktuell der vielversprechendste Weg. Jedoch angesichts der Bandbreite an ökonometrischen Methoden, die zur Messung ökonomischer Wirkungen zur Verfügung steht, merkt beispielsweise Godin (2006) kritisch an, dass für nicht-ökonomische Wirkungsdimensionen lediglich „case studies“ verwendet werden.

Um trotz dieser Vielzahl an Herausforderungen nicht-ökonomische Wirkungsdimensionen erfolgreich in Wirkungsanalysen zu implementieren, braucht es eine Reihe an veränderten und neuen Voraussetzungen.

Beschäftigt man sich mit sozio-ökonomischen Wirkungen in Evaluierungen, so braucht es eine Theorie, die die Fördermaßnahme mit den Daten und resultierenden Wirkungen verknüpft. Dies gilt natürlich auch für die Analyse ökonomischer Wirkungen. Es gibt zwar eine Vielzahl an impact – Indikatoren von FTI-Programmen, wie zum Beispiel „economic growth“, „productivity“, „profits“, „job creation“, „market share“ und „spin-offs“, jedoch nur eine geringe Anzahl von Indikatoren, die FTI-Fördermaßnahmen direkt mit ökonomischen „payoffs“ verknüpfen. Cozzens (2002) drückt dies folgendermaßen aus: „What the research evaluation community lacks (...) is a logic connecting concrete S&T policies and programs to the available data on outcomes“ (Cozzens 2002: 105).

Weiters braucht es neue Datensets. Um nicht-ökonomische Wirkungen umfassend analysieren zu können, wird es notwendig sein, ökonomische Datensets durch multidimensionale Datensets zu ersetzen, die soziale, kulturelle, gesundheitsspezifische und umweltbezogene Themen erheben. „Using existent data for new indicators is not a good idea, what we will get will be more of the same (Godin, 2006: 5)“. Des Weiteren sollten diese Datensets so ausgestattet sein, dass die Berechnung detaillierter nationaler und lokaler Ergebnisse möglich ist.

Zusätzlich muss die Erarbeitung neuer Indikatoren-Sets erwähnt werden. Nicht-ökonomische Effekte sind von Natur aus sehr komplex und erfordern daher ein spezifisches Messkonstrukt, das verschiedene Indikatoren zur Messung ein und desselben Konzepts vorsieht. Zeigen alle Indikatoren in dieselbe Richtung, so erhält man ein robusteres Ergebnis in Bezug auf die potentielle Wirkung einer Maßnahme, als wenn nur ein Indikator erhoben worden wäre. Gault (2011) verdeutlicht dies am Beispiel des Bruttonationalprodukts pro Kopf. „GDP is an indicator of the size of the economy and its change over time is an indicator of economic growth or decline. Neither are indicators of social progress or of wellbeing. GDP can be combined with other indicators, such as population, to give GDP/Capita. GDP/Capita is an example of an indicator that can mislead. It is used for ranking countries and the assumption is that a high ranking is good. However, two countries with a high ranking, such as Norway or the United States, could have quite different

income distributions (CIA 2010). The caveat that this illustrates is that no decision or policy should be based on just one indicator” (Gault, 2011: 3).

Abschließend braucht es neue Methoden. Es hat wenig Sinn, das bestehende „ökonomische System“ auf nicht-ökonomische Effekte anzuwenden. Godin (2002) deutet hierbei auf einen interessanten Aspekt hin: „A good example is innovation. Without really noticing that they had departed from their original goal, national governments and the OECD ended up measuring innovation the way they measured R&D in terms of input and activities” (ibid.: 24).

Nicht-ökonomische Effekte brauchen eine neue „Erhebungslogik“ – folglich ist die Entwicklung und Erprobung neuer sozialwissenschaftlicher Methoden notwendig.

6.4. WIRKUNGSDIMENSIONEN MESSBAR MACHEN – JENSEITS DER ÖKONOMISCHEN DIMENSION

Studien, in denen nicht-ökonomische Effekte gemessen werden, sollten folgende Aspekte beachten. Zum einen braucht es eine exakte Definition des „impacts“ und eine klare Zuordnung dieses Impacts zu einer eindeutig definierten/ klar umgrenzten Maßnahmen. „An impact is measured by a changed situation“ (Godin/Doré, 2006: 11).

Zum anderen sollten Veränderungen laut Godin/Doré auf zwei verschiedene Arten erhoben werden:

1. Als nominale Information (z.B. erhöhte sich die Anzahl an Publikationen? Ja/nein). Dieser Indikator macht es möglich, die Häufigkeit und die Art der Veränderung zu erfassen.
2. Als ordinale/numerische Information (z.B. wie viele Publikationen wurden veröffentlicht? Anzahl an Publikationen). Dieser Indikator zeigt die Bedeutung/Stärke der Veränderung an.

(Godin/Doré, 2011: 11)

Weiters sollten so genannte „questions on transfer“ überlegt und operationalisiert werden, die „input, output und impact“ miteinander verknüpfen. Fehlt der Impact einer Maßnahme, so bedeutet dies nicht, dass die Maßnahme/ das Projekt sinnlos war; es könnte einfach der „transfer“ nicht gelungen sein. Godin (2006) betont daher: „It is necessary therefore to question on the conditions, the context and the efforts of the organisation’s appropriation to appreciate the impact of science on it (Godin/Doré, 2006: 11)“.

Das „Linking“ zwischen „input und output“ ist fundamental, das betont auch Cozzens (2002): „We need to link back more to fundamental social problems and issues, rather than focusing narrowly on immediate payoffs from a particular program or activity“ (Cozzens, 2002: 75).

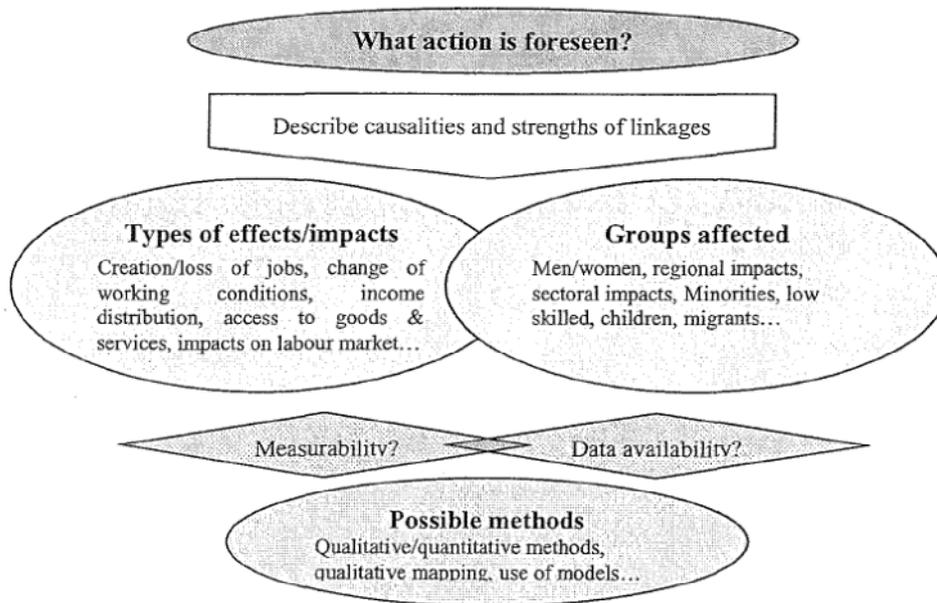
Die Europäische Kommission (2009) schlägt in ihrem Paper „Guidance for assessing Social Impacts within the Commission Impact Assessment system“ die Beantwortung folgender Fragestellungen bei der Evaluierung nicht-ökonomischer Wirkungen vor:

- Welche Maßnahmen werden durchgeführt?
- Welche Effekte sind zu erwarten?
- Auf welche Personengruppen wirken die Maßnahmen und auf welche Art und Weise?

- Welche Methoden können verwendet werden, um die entsprechenden Effekte zu quantifizieren?

Abbildung 36 stellt diesen Arbeitsprozess dar:

Abbildung 36: Arbeitsschritte einer Evaluierung



Europäische Kommission 2009

Da man es bei der Analyse nicht-ökonomischer Effekte häufig mit komplexen Auswirkungen auf die Gesellschaft zu tun hat, sollte hier insbesondere Punkt 2 und 3 (Welche Effekte sind zu erwarten? Auf welche Personengruppen wirken die Maßnahmen und auf welche Art und Weise?) besonders beachtet werden.

Fördermaßnahmen können unterschiedliche Effekte auf unterschiedliche Personengruppen haben. Deshalb empfiehlt es, sich Personengruppen entlang von sozio-ökonomischen Charakteristika zu bilden und getrennt zu analysieren. Eine „Gender-Perspektive“ sollte standardmäßig Teil der Programmlogik sein. Des Weiteren gilt es, die Frage nach der Art der Effekte „offen“ zu stellen und sowohl positive als auch mögliche negative Effekte ins Blickfeld zu nehmen. Eine Wirkungsanalyse sollte ambivalente Effekte sichtbar machen.

Bei der Auswahl der passenden Indikatoren sollten folgende Aspekte beachtet werden:

Ein Indikator dient ausschließlich als ein „Anzeiger“ für untersuchte Sachverhalte und Prozesse. Die Europäische Kommission (2009) definiert Indikatoren folgendermaßen: „An indicator is a quantitative or qualitative factor of variable that provides a simple means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor“ (Garbarino/Holland, 2009: 6)

Ein Indikator gestattet nur, mit einer gewissen Wahrscheinlichkeit auf das vorausgesetzte Indikaturniveau zu schließen. Bei komplexen Sachverhalten sollten daher mehrere Indikatoren eingesetzt werden, um eine Eigenschaft eines Untersuchungsobjekts zu repräsentieren.

Indikatoren sind zumeist nicht „universell“ anwendbar, sondern müssen in Bezug auf folgende Kontexte adaptiert und kontinuierlich in Bezug auf ihre statische Validität und Reliabilität getestet werden:

- nationaler /lokaler Kontext
- wissenschaftliche Disziplin
- nationale Sprache
- Individualebene/aggregierte Ebene

Laudel /Gläser (2011) zeigen am Beispiel der „wissenschaftliche Publikationen“ sehr anschaulich, welche Schwierigkeiten entstehen, wenn ein Indikator, der auf der aggregierten Ebene seinen Zweck erfüllt, auf die Individualebene angewendet wird.

„Publication indicators offer only partial insight into research quality at best, and are usually unreliable, when applied to individual researcher and their research“ (Laudel/Gläser, 2011: 5). Auf der Individualebene stellt die Anzahl an Publikationen keinen validen Indikator für „research performance“ dar; dieser Indikator zeigt lediglich die Forschungsaktivität eines/einer Wissenschaftlers/Wissenschaftlerin an. „The numbers of publications offer little useful information. They are even dangerous because their use can incite mass production of useless publications“ (ibid. 8). Weitere Indikatoren, die häufig gewählt werden, um wissenschaftliche Exzellenz anzuzeigen, sind Preise, Auszeichnungen oder Einladungen zu Vorträgen. Laut Laudel/Gläser zeigen diese Indikatoren die symbolische Bewertung an, die die Forschungscommunity den einzelnen Forschenden beimisst. Die Validität dieser Indikatoren ist zu hinterfragen, da die Preise und Auszeichnungen zumeist idiosynkratisch und daher nicht vergleichbar sind und die Auswahl der ReviewerInnen oft unter großem Zeitdruck erfolgt.

Laudel/Gläser (2011) schlussfolgern daher „Indicators which work at the aggregate level, lose their power if applied to individuals, in order to learn something about an individual case, we need to study in depth“ (Laudel/Gläser, 2011: 9). Die AutorInnen empfehlen das Augenmerk auf prozessorientierte Indikatoren zu legen, z.B. mit Reviewern und dem/der WissenschaftlerIn selbst die Dynamik der Forschungskarriere zu reflektieren. Die einzelnen Forschungsphasen, die WissenschaftlerInnen durchlaufen, können oftmals an der Reihenfolge der AutorInnen auf Publikationen nachvollzogen werden. Dazu Laudel/Gläser (2011): „According to this rule, the researcher who conducted the experiments becomes first author, the head of the laboratory (who often defined the problem and contributed theoretical background) is last author, and other collaborators are listed in between. In fields where this rule applies, the progress of a researcher’s career is reflected by the move of his or her name from first author to intermediate author to last author“ (ibid. 11).

Box 20: Beispiele für Indikatoren zur Bestimmung wissenschaftlicher Qualität auf der Individualebene

Weiters eignen sich insbesondere die folgenden drei Indikatoren, die „wissenschaftliche Karriere von ForscherInnen auf der Individualebene zu bestimmen):

- „Influential research contributions“: The influential research contributions often consist of linking fields and ideas that their community thought of as separate. By creating such new, interdisciplinary links, elite researchers often open-up whole new areas of research opportunities for one or more communities.
- „Long-Term research programmes“: Elite researchers have long-term programmes.
- „Collaboration networks“: Scientists always try to find the best experts on this knowledge and to collaborate with them. Collaboration networks are stratified according to the quality of collaborators. This means that researchers who collaborate with the best groups in a field are likely to belong to the elite themselves.

(Laudel/Gläser, 2011: 11)

Als nächster Punkt soll der Einsatz von so genannten „composite indicators“ (deutsch: zusammengesetzte Indikatoren) behandelt werden.

Zusammengesetzte Indikatoren basieren auf aggregierten und gewichteten Einzelindikatoren. Sie ermöglichen die Operationalisierung komplexer und multidimensionaler Themenbereichen, wie z.B. Umwelt, Gesundheit sowie soziale und technologische Entwicklung, können aber auch „in die Irre“ leiten, sofern ihre Konstruktion methodisch problematisch und die Interpretation schwierig ist.

Werden zusammengesetzte Indikatoren im FTI-Bereich auf der aggregierten Ebene verwendet, so müssen stets die nationalen/lokalen Kontexte sowie im Falle von Publikationen die Landessprache und die wissenschaftliche Disziplin mitbedacht werden.

Wir empfehlen einen eher vorsichtigen Umgang mit zusammengesetzten Indikatoren für nicht-ökonomische Wirkungsdimensionen auf der Ebene einzelner Unternehmen.

Laudel/Gläser zeigen in ihrer Studie über den Hirsch Index (Einzelindikatoren: Anzahl an Publikationen, Anzahl an Zitationen), der den wissenschaftlichen Einfluss eines/einer ForscherIn abbilden soll, dass WissenschaftlerInnen, die viel publizieren, aber selbst kaum zitiert werden, und WissenschaftlerInnen, die wenige Publikationen veröffentlichen, die aber äußerst erfolgreich zitiert werden, denselben Hirsch-Index Wert zugewiesen bekommen.

Ebenso kommen Warren et al. (1998) zu dem Schluss, auf zusammengesetzte Indikatoren auf der individuellen Ebene zu verzichten, wenn Geschlechterunterschiede im Arbeitsmarkterfolg (Einzelindikatoren: Einkommen, Bildung, Beruf) dargestellt werden. Sie begründen dies wie folgt: „First, women often have higher levels of education than men in the same occupation, while men usually have higher earnings than women in the same occupation. Thus, when a global index is

used, the relative standing of men's and women's occupations is an arbitrary function of the weights given to occupational education and occupational earnings" (Warren et al., 1998:5).

Geht es darum, komplexe, multidimensionale Wirkungen auf der individuellen Ebene zu operationalisieren, so sollten mehrere Indikatoren zum Einsatz kommen, welche die Effekte einzeln und unter mehrmaliger Wiederholung messen, um robuste und reliable Schätzwerte zu liefern.

6.5. MESSUNG NICHT-ÖKONOMISCHER WIRKUNGSDIMENSIONEN

In Bezug auf Gesundheits- und Umweltaspekte werden häufig „metrics of social impacts“ berechnet, da der kausale Zusammenhang zwischen Investment und Impact klarer ersichtlich ist als bei anderen nicht-ökonomischen Wirkungsdimensionen, wie z.B. bei der Messung des Sicherheitsgefühls einer Bevölkerung oder dem Konzept der Lebensqualität.

Nicht-ökonomische Effekte können dann am einfachsten „finanziell“ bewertet werden, wenn ihr „Wert“ einem „Marktwert“ entspricht. Dies verdeutlicht folgendes Beispiel der Europäischen Kommission: „Air pollution damage to crops might reduce crop yields, thus allowing for relatively straight forward monetisation" (IA EC, 2009: 40 – Part III, Annexes)

Können nicht-ökonomische Wirkungsdimensionen nicht direkt in Marktpreisen ausgedrückt werden, stehen folgende Methoden zur Verfügung:

- „Stated preference methods – Präferenzen werden erfragt (Verfahren: contingent valuation, conjoint analysis, choice experiments)“ und
- „Revealed preference methods – Präferenzen werden aus Preisen abgeleitet (Verfahren: travel cost method, hedonic pricing)“.

Diese Methoden haben folgendes Ziel: „They are ways of calculating monetary costs and benefits of goods that do not have a direct market price" (IA EC, 2009: 40 – Part III, Annexes).

Stated Preference Methoden basieren auf Befragungen von Personen über ihre individuelle Zahlungsbereitschaft bzw. Akzeptanzbereitschaft sowie ihre Eigenschaften, Haltungen und Einstellungen, aber auch ihr mögliches Verhalten in Bezug auf hypothetische Veränderungen.

Mit Hilfe dieses Instruments können mit vergleichsweise geringem Aufwand Marktchancen bewertet, Potenziale identifiziert und konkrete Aussagen zur Optimierung neuer Produkte oder Dienstleistungen getroffen werden.

Die Auswertung von Stated-Preferences-Daten erlaubt es, Antworten zu beispielsweise folgenden Fragestellungen zu finden:

- Welche Merkmale des Angebots haben für die Erhöhung der Nachfrage die größte Bedeutung?
- Wie ändert sich die Nachfrage, wenn wir den Preis für Produkt X um x% erhöhen?
- Was passiert, wenn wir eine ganz neue Dienstleistung oder ein neues Verkehrsmittel einführen?
- Wie kann das vorhandene Angebot am effektivsten verändert werden?

Bei Revealed Preference Methoden wird der Wert von z.B. Umweltgütern aus dem Markt für verbundene Güter (z.B. Wohnen) abgeleitet. Diese Methode basiert auf dem tatsächlichen Verhalten von AkteurInnen.

Nicht-ökonomische Wirkungsdimensionen, die auf verhaltensbezogene Änderungen abzielen (Veränderungen der Werte, des Images, etc.), aber auch komplexe Wirkungstypen wie Inklusion, Diversität, Gleichberechtigung, sozialer Zusammenhalt oder soziale Innovation können durch folgende Attribute bestimmt werden: sie sind dynamisch, komplex, zum Teil „global“ und nicht-linear (Gault 2011).

Für eine große Anzahl an nicht-ökonomischen Wirkungsdimensionen kommen aufgrund dieser Attribute sozialwissenschaftliche Methoden als Erhebungs- und Auswertungsinstrumente in Frage, ökonometrische Methoden treten hier in den Hintergrund.

Die OECD (2008) drückt dies folgendermaßen aus: „In general, it must be realised that the most that can be done is to highlight where these impacts occur and articulate qualitatively the <value> of these impacts on society“ (OECD, 2008: 213).

Case studies, ExpertInnenbefragungen und qualitative Interviews sind vielversprechende Methoden „to tell the story of the impacts“ (OECD, 2008: 213).

Des Weiteren zeigt sich international der Trend, insbesondere bei der Evaluierung nicht-ökonomischer Wirkungsdimensionen, partizipative Methoden¹⁶ zu verwenden.

Die Stärke von partizipativen Methoden ist es, die „Öffentlichkeit“, relevante „stakeholder“ oder „enduser“ direkt in den Evaluationsprozess miteinzubeziehen. Auf Basis dieser Methoden können sich auch nicht-wissenschaftliche Zielgruppen intensiv in das Programm-Projektgeschehen integrieren und ihr Wissen sowie ihre Kenntnisse, Ansichten, Werte und Argumente ausdrücken und austauschen. Dabei kommen qualitative sowie quantitative partizipative Methoden (z.B.: participatory social mapping, causal-linkage, change and trend diagramming) zum Einsatz. Auf diese Weise können für die Betroffenen relevante Evaluierungsfragen gestellt und auf die Zielgruppe abgestimmte Evaluierungsmethoden eingesetzt werden. Weiters kann auch im Kontext politischer Entscheidungen und Maßnahmenplanungen das Wissen und der Sachverstand der relevanten Interessensgruppierungen berücksichtigt werden; beides ist bei der Formulierung von Lösungsstrategien von entscheidender Bedeutung.

¹⁶ Für einen Überblick über partizipative Methoden und ein partizipatives Evaluationsdesign, siehe: http://www.rat-fte.at/tl_files/uploads/Studien/060201_Leitfaden%20partizipativer%20Verfahren_ITA.pdf

7 Datenverfügbarkeit als kritischer Engpassfaktor bei Evaluierungen

“The errors which arise from the absence of facts are more numerous and durable than those which result from unsound reasoning respecting true data.”

Charles Babbage, zitiert nach Rosenberg (1994:27)

Die Datenverfügbarkeit ist die entscheidende Determinante der Methodenwahl in der Evaluierungspraxis. Die Datenlage ist primär aus der Perspektive quantitativer Methoden von Bedeutung. Dies lässt sich anhand von drei Argumenten begründen: Erstens sind für qualitative Untersuchungen keine Sekundärdaten vorhanden. Zweitens wird im Rahmen von Evaluierungen eine Befragung einer geringen Anzahl von Programmteilnehmern mithilfe qualitativer Interviews als weniger problematisch betrachtet wie eine Befragung einer größeren Stichprobe von Unternehmen. Letzteres ist aufgrund verschiedener Ursachen (Bürokratiekosten, Lobbying, etc.) zunehmend unerwünscht. Drittens sind qualitative Forschungsdesigns regelhaft nicht an die Verfügbarkeit einer Kontrollgruppe gebunden. Demnach sind die Aussagen folgenden Unterkapitel primär auf ökonometrische Methoden anzuwenden.

7.1. DIE KOMPLEMENTARITÄT VON DATEN UND METHODEN – EINE PRODUKTIONSTHEORETISCHE PERSPEKTIVE

Auch wenn die Anzahl der Förderfälle, Programmdesign und die Programmziele eine analytische, quantitative Methode nahelegen, kann diese nicht angewandt werden wenn die entsprechenden Daten fehlen. Es nützt gar nichts, Vorschläge über mögliche Methodenanwendungen bei unterschiedlichen Programmen zu machen, so lange die Datenlage dies nicht ermöglicht. Daher kommt aus der Perspektive der Studienautoren dieser Frage eine entscheidende Bedeutung zu.

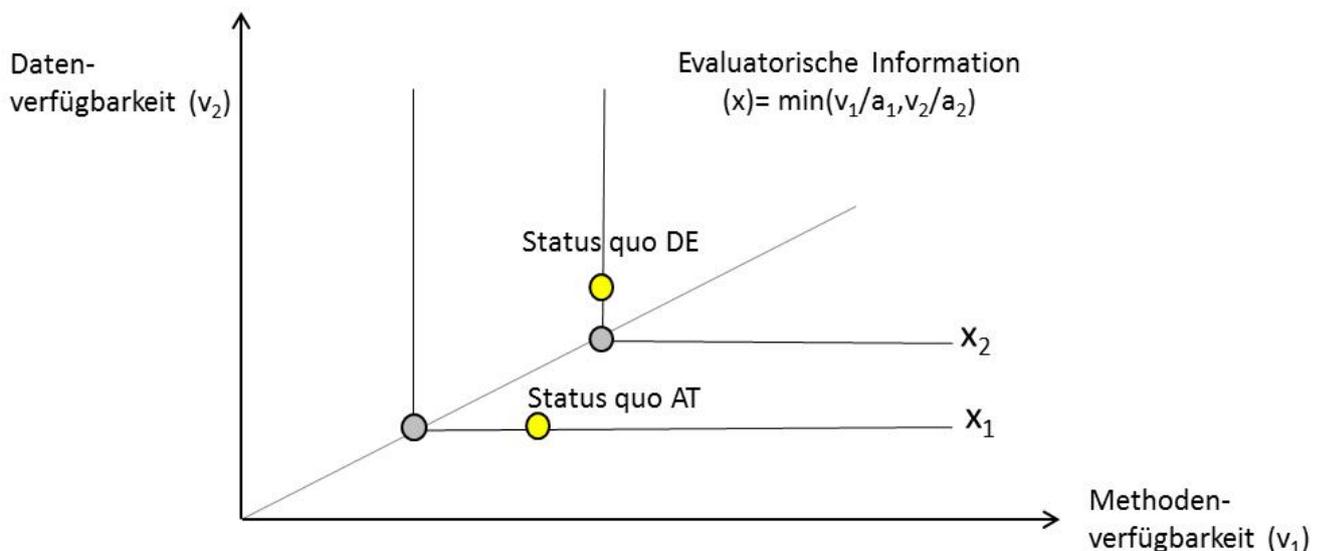
Die methodischen Entwicklungen und Fortschritte im Bereich der Mikroökonomie etwa zeigen, dass diese auch eine Funktion verbesserter Datenlage seit den 1960er Jahren sind (Angrist und Pischke 2010; The Royal Swedish Academy of Sciences 2000). In diesem Sinne kann eine substantielle Weiterentwicklung von quantitativen FTI-Evaluationsmethoden bzw. die Adaption innovativer Methoden aus anderen Bereichen nur bei verbesserter Datenqualität stattfinden. Dieser enge Konnex von Daten und Methoden lässt sich anhand einer Leontief-Produktionsfunktion konzeptualisieren. Dieser Funktionstyp zeichnet sich dadurch aus, dass die Produktionsfaktoren nicht substituierbar sind. Das bedeutet, dass eine erhöhte Methodenverfügbarkeit, schlechte Daten nicht oder jedenfalls nur in einem geringen Ausmaß ersetzen kann. Diese Feststellung ist freilich nicht unumschränkt gültig, denn tatsächlich wurden einige methodische Weiterentwicklungen auch aus einem Engpass auf der Datenseite heraus entwickelt. Aber nach Heckman et.al. (1999:1868) hat auch diese Strategie mittlerweile das Stadium deutlich abnehmender Grenzerträge erreicht: „Too much has been asked of econometric

methods to remedy the defects of the underlying data. When certain features of the data are improved, the evaluation problem becomes much easier.

The best solution lies in improving the quality of the data on which evaluations are conducted and not in the development of formal econometric methods to circumvent inadequate data.”

Für die FTI-Evaluierung dürfte die Beschreibung durch eine limitationale Produktionsfunktion aber sehr wohl zutreffend sein. Insbesondere gilt hier für die österreichische Situation, dass Daten über Kontrollgruppen schlichtweg fehlen. Hier ist tatsächlich die Nichtverfügbarkeit und nicht mangelnde Qualität das Problem. Abbildung 37 zeigt schematisch einen Produktionsprozess evolutorischer Information mit einer Leontief-Technologieannahme: Die beiden Produktionsfaktoren sind Methoden- und Datenverfügbarkeit; der Produktionsmöglichkeitsraum besteht aus unterschiedlichen Mengen evaluatorischer Information. Betrachtet man etwa die Isoquante X_1 , so zeigt sich, dass ausgehend von der Ecklösung keine Zunahme an evaluatorischer Information erzielt werden kann, wenn der jeweils andere Faktor konstant bleibt. Diese Situation passt relativ gut für Österreich: Während mittlerweile zahlreiche Methodenstudien und Methodensammlungen – wie auch die hier vorliegende – auf die Möglichkeit und Sinnhaftigkeit eines elaborierter Methodeneinsatzes hinweisen, hat sich hinsichtlich Datenverfügbarkeit nichts verändert. Entsprechend symbolisiert der Punkt AT die Situation Österreichs: Schlechte Datenlage verhindert den Einsatz innovativer, quantitativer Methoden. Damit werden wesentliche Potentiale zur Realisierung einer „evidenced-based policy“ nicht ausgeschöpft und das Wissen über einen effektiven und effizienten Einsatz öffentlicher Ressourcen für F&E bleibt limitiert. Die in Box 2 (Kapitel 2) formulierten FTI-politischen Ziele zur Weiterentwicklung von Evaluierungsmethoden können so wohl kaum erfüllt werden.

Abbildung 37: Evaluatorische Information wird mittels einer limitationalen Produktionsfunktion generiert



Eigene Abbildung

Abbildung 37 stellt weiterhin einen schematischen, und im Einzelfall freilich nicht notwendigerweise zutreffenden Vergleich Österreichs mit der deutschen Situation dar. Demnach befindet sich Deutschland – jedenfalls was die durch quantitative Methode gewonnene evaluatorische Information im FTI-Bereich anlangt – auf der höheren Isoquante X_2 . Generell lässt sich feststellen, dass Deutschland durch eine höhere Methodenkompetenz und eine bessere Datenverfügbarkeit gekennzeichnet ist. Dieser Befund wird insbesondere durch die Aktivitäten des Zentrums für Europäische Wirtschaftsforschung (ZEW) sowie das Rheinisch-Westfälischen Instituts für Wirtschaftsforschung (RWI Essen) untermauert. Wenngleich freilich auch in Deutschland die Datenlage besser sein könnte (wo nicht?), so kann doch die Hypothese aufgestellt werden, dass hier möglicherweise eher die Methodenverfügbarkeit als limitierender Faktor wirkt.

Die vermutlich höhere Methodenkompetenz in Deutschland ist dabei auch ein Ergebnis einer bereits längeren Tradition in der Anwendung ökonomischer Methoden auf evaluatorische Fragestellungen im FTI-Bereich. Dabei sind zwei Effekte zu unterscheiden. Zunächst einmal erfordert die Arbeit mit einem neuen Datensatz, z.B. der F&E-Erhebung, eine Aufwendung in Form von teilweise relativ hohen Fixkosten, weil diese keineswegs stets offensichtlichen Besonderheiten eines Datensatzes berücksichtigt werden müssen. Zweitens ist zu bedenken, dass quantitative, empirische Wissenschaft kein isoliertes Projekt darstellt. Vielmehr entsteht durch die hohe Standardisierung und die explizite Darstellung der Methode sowie der Daten die Möglichkeit zu einem kumulativen Wissensaufbau: „In the empirical sciences, evidence accumulates across settings and study designs, ultimately producing some kind of consensus. Small ball sometimes wins big games.“ (Angrist und Pischke 2010:25; siehe hierzu auch Heckman und Smith 1995). Kurz gesagt: Es besteht in Deutschland ein Erfahrungsvorsprung in der Analyse quantitativer Daten, der positiv auf die Methodenkompetenzen zurückwirkt. Dieser Aspekt eines kumulativen Wissensaufbaus über die Wirkungen von Programmen ist ein entscheidender Aspekt empirischer Politikevaluation mittels ökonomischer Methoden.

7.2. DIE DATENLAGE IN ÖSTERREICH

Nachdem bereits in den vorangehenden Kapiteln auf die relativ inferiore Datenlage in Österreich Bezug genommen wurde, soll diese hier etwas detaillierter dargestellt werden. Freilich ist diese Feststellung keineswegs neu. So haben etwa Leo et al. (2006:39) im WIFO-Weißbuch bereits auf die Problematik hingewiesen: „Dass solche Analysen [über die Effektivität von Förderungen hinsichtlich unterschiedlicher Outputindikatoren, Anm. d. Verf.] weitgehend fehlen, liegt an der unzureichenden Verfügbarkeit entsprechender Datensätze.“ Auch Rhomberg et al. (2006) stellen im gleichen Jahr fest, dass die niedrige Fallzahl ökonomischer Evaluierungen primär eine Folge schlechter Datenlage ist.

Zunächst soll anhand des Beispiels der Datenlage zur indirekten Forschungsförderung eine Illustration des Status quo gegeben werden. Dieses Beispiel eignet sich aufgrund seiner empirischen Bedeutsamkeit besonders gut um die Relevanz von Datenverfügbarkeiten zu erörtern. Es wurde bereits die Asymmetrie in evaluatorischer Hinsicht zwischen direkter und indirekter Forschungsförderung festgestellt. Um das folgende Argument nochmals empirisch zu fundieren,

zeigt Tabelle 23 die Ergebnisse über die öffentliche Finanzierung von F&E im Unternehmenssektor der aktuellen F&E-Erhebung. Dabei ist zu beachten, dass nach dem Frascati-Manual nur die nicht-rückzahlbaren Zuschüsse der Rubrik FFG zugerechnet werden, während rückzahlende, zinsbegünstigte Darlehen den unternehmenseigenen Mitteln zugerechnet werden. Demnach entfielen auf die Forschungsprämie mit 246 Mio. € beinahe 60% öffentlicher F&E-Subventionen für Unternehmen.

Tabelle 23: Öffentliche Finanzierung der Ausgaben für F&E im Unternehmenssektor 2009
(Firmeneigener Bereich)

	Bund	Forschungsprämie	Länder	FFG ¹	Sonst. öffentl. Finanzierung
Absolut (in 1.000 €)	18.499	246.622	18.190	127.869	8.307
% öff. Finanzierung	4.4	58.8	4.3	30.5	2.0

Schiefer 2011; 1: Nur Zuschüsse, keine zinsbegünstigten Darlehen

Vor diesem quantitativen Hintergrund ist das folgende Zitat aus der Systemevaluierung zu lesen. Es zeigt eindrucksvoll das Mismatch hinsichtlich Erwartungen an Datenverfügbarkeit zwischen EvaluatorenInnen und Politik bzw. Agenturen in Österreich:

„The statistical basis for a thorough evaluation of Austrian tax funding turned out to be insufficient in many respects [Herv. d. Verf.]. There are severe deficiencies in the coverage of relevant information on the one hand, and abundant amounts of very detailed but useless information on the other. In parts, the statistics misrepresent evidence on the use of tax incentives. Access to a unique and potentially very valuable firm-level database on the usage of allowances for inventions was denied since data privacy laws apply. The Ministries of Finance and Economics eventually provided aggregate data [und eben keine Mikrodaten, wie dies aber für international übliche ökonometrische Verfahren (e.g. Berube und Mohnen 2007) notwendig wäre!, d. Verf.]. Empirical evidence based on micro-data would have been much more reliable and explicit. Above all, the search for data tied up far too many resources and could hardly have been accomplished by non Austria-based evaluators. To the extent that Austrian policy makers value quantitative assessments of tax funding measures, possibly commissioned to international experts, such data should in future be made readily available.

Insisting on data confidentiality produces some unpleasant side effects. Policy making becomes less evidence based and more subject to the interest and judgments of selected opinion leaders. At best, these interests are transparent (“Standort-Additionalität”); at worst, they enter the policy discussion in a very subtle way and become hard to grasp.

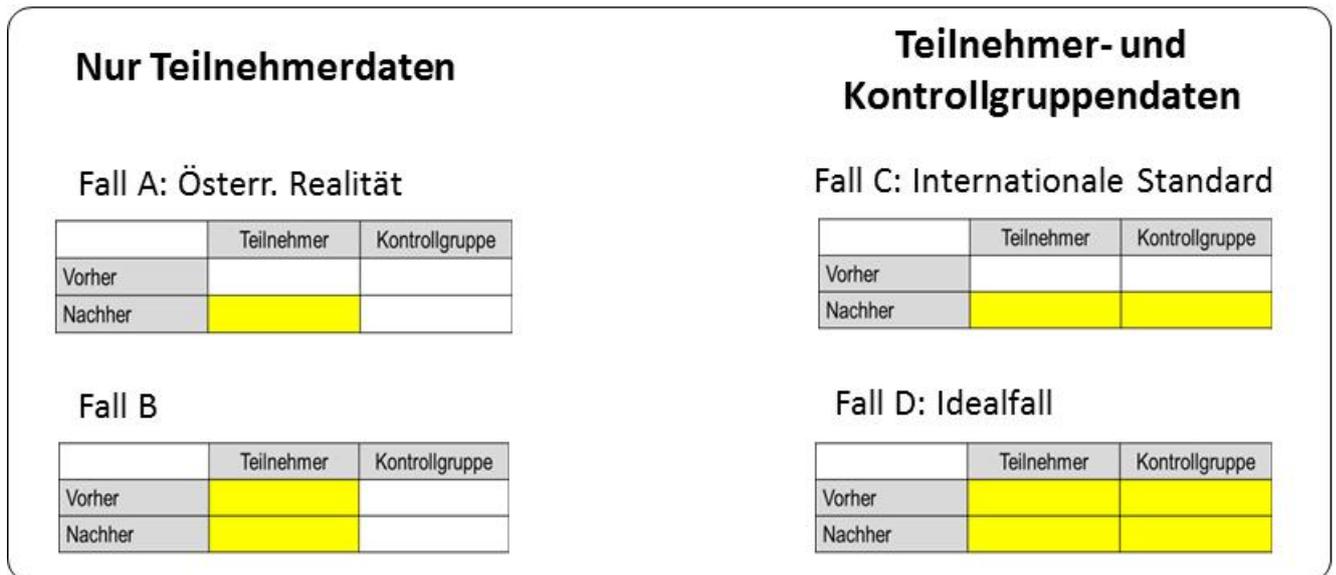
In any case, distinguished international academic scholars (from KU Leuven, from Merit in Maastricht) refrained from participating in the evaluation of R&D tax incentives when they were invited to do so in late 2007. In their opinion, this task could not be accomplished with the poor database at hand.”

(Falk et al. 2009: 7f, 33.)

Erschwerend kommt hinzu, dass diese Problematik bereits in einem Gutachten des Rechnungshofs von 2007 angemerkt wurde: „In Österreich wurde seit dem Jahr 2000 die indirekte (steuerliche) Forschungsförderung zur Hebung der Forschungsquote kontinuierlich ausgebaut. Die Wirkungen der indirekten Forschungsförderung wurden jedoch bisher noch nicht evaluiert.“ Obwohl bislang aufgrund mangelnder Datenverfügbarkeit noch immer keine Evaluierung nach internationalen und methodisch anspruchsvollen Standards stattgefunden hat (Janger et al. 2010), wurde die Forschungsprämie von 8 auf 10% erhöht und damit die Bedeutung der indirekten Forschungsförderung weiter ausgebaut. Die durch das Finanz- und Wirtschaftsministerium bereitgestellten, unzureichenden Daten sind die entscheidende Ursache für diese negativ zu beurteilende Konstellation. Warum das Finanzministerium im Rahmen der Wirkungsorientierten Haushaltsführung einerseits eine Kontrolle der Wirkungen öffentlichen Ressourceneinsatzes verlangt, aber im eigenen Verantwortungsbereich nicht die notwendigen Daten zur Verfügung stellt um diesem Ziel zu entsprechen, muss wohl eine offene Frage bleiben.

Um nun die Diskussion ein Stück weit zu verallgemeinern und um auch die Konsequenzen unterschiedlicher Datensets auf mögliche Identifikationsannahmen zu untersuchen, stellt Abbildung 38 einen systematischen Überblick zu verschiedenen Datenlagen dar. Zunächst erfolgt hierbei eine Einteilung in Datensets die nur Daten über geförderte Unternehmen beinhalten und Datensets die auch Beobachtungen zu nicht-geförderten Unternehmen inkludieren. Die zweite Eigenschaft anhand derer Datensets für FTI-Evaluierungen differenziert werden können, ist die zeitliche Dimension. Liegen Daten nur zu einem Zeitpunkt t nach der Förderung vor, oder auch zu einem Zeitpunkt vor der Förderung? Diese zwei Faktoren ergeben in Summe vier Fälle von Daten (A bis D), wobei diese Fälle aus methodischer Perspektive ordinal skaliert sind und Fall A die schlechteste und Fall D die beste Datenlage repräsentieren. Im engeren Sinne ermöglichen nur Datensätze in der rechten Spalte von Abbildung 38 eine Orientierung an modernen, quantitativen, analytischen Methoden: „(...) this literature [on program evaluation with microeconomic methods, [Anm. d. Verf.] is focused on settings with observations on units exposed and not exposed, to the treatment, with the evaluation based on comparisons of units exposed and not exposed.“ (Wooldridge 2009:6). Nun zu den einzelnen Fällen und den damit möglichen ökonomischen Analysemethoden.

Abbildung 38: Datensätze zur Evaluierung FTI-politischer Maßnahmen



Eigene Darstellung

- **Fall A:** Hier liegen nur Daten über die Teilnehmer nach Maßnahmenteilnahme vor. Diese Datenkonstellation trifft für die absolute Mehrzahl der österreichischen FTI-Evaluierungen zu. Es muss jedoch festgestellt werden, dass dies gleichzeitig die schlechteste aller denkbaren Datensituationen darstellt. Die direkte Frage nach den Programmwirkungen beim geförderten mag zwar zunächst als billige und naheliegende Evaluierungsmethode erscheinen. Sie ermöglicht aber letztlich keine methodisch nachvollziehbare Identifikation kausaler Effekte und ist von zahlreichen Problemen gekennzeichnet. Letztlich sind damit alle Ansätze auf Basis von Daten nach dem Fall A theoretisch nicht fundierte Kausalanalysen.
- **Fall B:** Hier liegen Daten zur Performance der Unternehmen vor und nach Maßnahmenteilnahme vor. Liegen mehrere Beobachtungen über die Zeit für die gleichen Unternehmen vor, spricht man von Paneldaten. Damit ist die Identifikationsannahme eines Vorher-Nachher-Schätzers möglich. Auch diese – relativ zu Fall A – bereits etwas bessere Datenlage liegt bei österreichischen FTI-Evaluierungen kaum vor. Einzige Ausnahme stellt das Programm KIRAS dar. Allerdings ist auch diese Identifikationsannahme keineswegs unproblematisch und Czarnitki et al. (2003) raten von einer Anwendung von Vorher-Nachher-Schätzern zur Identifikation kausaler Effekte von FTI-Programmen ab.
- **Fall C:** Die Beobachtung zu Teilnehmern und Nicht-Teilnehmern nach Maßnahmenteilnahme stellt die klassische Datenlage für ökonometrische Kausalmodelle dar. Die damit ermöglichten Querschnittsverfahren können durch unterschiedliche Methoden für systematische Unterschiede zwischen den Teilnehmern und Nicht-Teilnehmern kontrollieren. Damit werden der Vergleich zwischen der Performance beider Gruppen valide und eine kausale Interpretation möglich. Allerdings ist hier auch anzumerken, dass die beiden klassischen Methoden, Matching und Kontrollfunktionsansätze, in einem ersten Schritt eine Modellierung der Teilnahmeentscheidung erfordern. Dafür sind unter Umständen auch Daten über Variablenwerte zum Zeitpunkt vor der Förderzusage notwendig. In Österreich liegt mit Polt

und Pointner (2005) bislang erst eine Studie vor, die mikroökonomische Modelle auf einen Querschnittsdatensatz angewandt hat vor. Auch Falk (2006) analysiert Daten zu geförderten und nicht geförderten Unternehmen, wobei letztere zwar um Förderung ansuchten aber keine zugeteilt bekamen. Allerdings wird hier keine ökonomische Kausalanalyse vorgenommen.

- **Fall D:** Auch hier liegen wie im Fall B Paneldaten vor, allerdings umfasst das Panel nunmehr Daten zu geförderten und nicht geförderten Unternehmen. Damit sind zunächst einmal alle Schätzer die auf Querschnittsdaten aufbauen möglich und zusätzlich vor allem auch Differenz-von-Differenzen-Schätzer. Diese nutzen die Datenlage in Fall D in optimaler Art und Weise aus. Solche Daten wurden in Österreich noch nie für FTI-Evaluierungszwecke analysiert.

Zusammenfassend kann damit festgestellt werden, dass die österreichische Datenlage durchwegs mangelhaft in Bezug auf ökonomische Analysen ist. Es gibt mit Polt und Pointner (2005) genau genommen erst eine ernst zu nehmende Evaluation auf Basis moderner ökonomischer Methoden, die internationalen Standards entspricht. Wenn sich an der Datenverfügbarkeit nichts verändert, wird es bis auf weiteres wohl auch bei diesem Befund bleiben. Der folgende Abschnitt zeigt anhand eines internationalen Vergleichs, dass es auch anders geht.

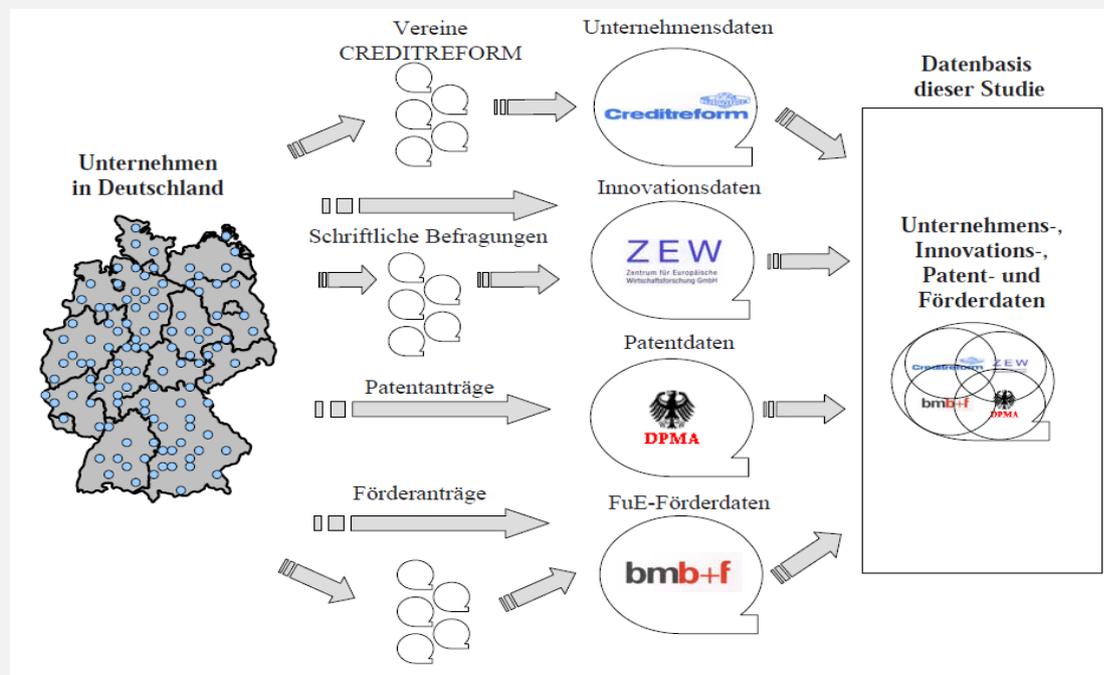
7.3. ES GEHT AUCH ANDERS – INTERNATIONALE VERGLEICHSBEISPIELE

Datenprobleme im FTI-Bereich sind keineswegs außergewöhnlich. So bemerkt etwa die Europäische Kommission in ihrer Kommunikation zur „Europe 2020 Flagship Initiative Innovation Union“ „that there is an urgent need to improve data availability and the breadth and quality of indicators to measure and monitor innovation performance (...)“ (EC 2010: 29). Trotzdem zeigen die folgenden Beispiele aus typischen Vergleichsstaaten, dass die Datenlage in Österreich zur Evaluation von FTI-Programmen im internationalen Vergleich signifikant schlechter ist. Dabei wird deutlich, dass nicht nur die skandinavischen Staaten eine bessere Verwertung ihrer administrativen Daten für Evaluierungszwecke aufweisen, sondern auch Deutschland und die Schweiz liberalere Normen (formell und/oder informell?) bezüglich Datenschutz als Österreich haben. Hier liegt jedoch auch insofern eine andere Konstellation als in Österreich vor, als wie dort nicht die statistischen Bundesämter die Innovationserhebung (Community Innovation Survey) durchführen, sondern Wirtschaftsforschungsinstitute. Diese können die Daten dann in Form von Mikrodaten weiter verwenden und auch mit anderen Datenquellen fusionieren bzw. „matchen“. In Deutschland trifft dies auf das Zentrum für Europäische Wirtschaftsforschung (ZEW) und in der Schweiz für die Konjunkturforschungsstelle der ETH Zürich (KOF) zu. Im Folgenden werden Beispiele für verwendete Datensätze für FTI-Evaluationen aus Deutschland, Schweiz, Flandern, Finnland und Dänemark dargestellt. Es sei an dieser Stelle aber betont, dass es auch Beispiele aus weiteren Staaten gibt, wo amtliche F&E-Statistiken mit weiteren Datensets auf Unternehmensebene verknüpft wurden um die Evaluation von F&E-Subventionen zu optimieren. Duguet (2004) evaluiert z.B. französische FTI-Politik mit französischen CIS-Daten, während Clausen (2007) norwegische CIS-Daten als Basis für eine ökonomische Mikrodatenanalyse heranzieht.

Box 21: Fallbeispiel Deutschland

Czarnitzki et al. (2003) haben für eine Studie im Auftrag des deutschen Bundesministeriums für Bildung und Forschung zum Thema „Einsatzmöglichkeiten neuer quantitativer Ansätze zur Wirkungsanalyse öffentlicher Forschungsförderung“ einen Datensatz erstellt, der auf der Fusion mehrere bestehender Sekundärdatensätze beruht. Wie Abbildung 39 zeigt, handelt es sich dabei um insgesamt vier verschiedene Datenquellen. Unternehmensdaten wurden auf der Basis einer Datenbank der Creditreform erhoben. Innovationsdaten stammen aus dem Mannheimer Innovationspanel (MIP), dem deutschen CIS. Über den Zugang zu den MIP-Daten heißt es auf der Homepage des ZEW: „Das ZEW legt Wert darauf, dass das MIP für wissenschaftliche Analysen zur Verfügung steht. Die Daten werden in anonymisierter Form (Scientific Use File) externen Nutzern für rein wissenschaftliche, nicht-kommerzielle Zwecke zur Verfügung gestellt. Zurzeit nutzen mehr als 30 Wissenschaftler die Scientific Use Files.“¹⁷ Weiterhin wurden Patentdaten des Deutschen Patentamtes sowie Förderdaten vom Bundesministerium für Bildung und Forschung verwendet.

Abbildung 39: Datenstruktur bei Czarnitzki et al. (2003)



Czarnitzki et al. 2003

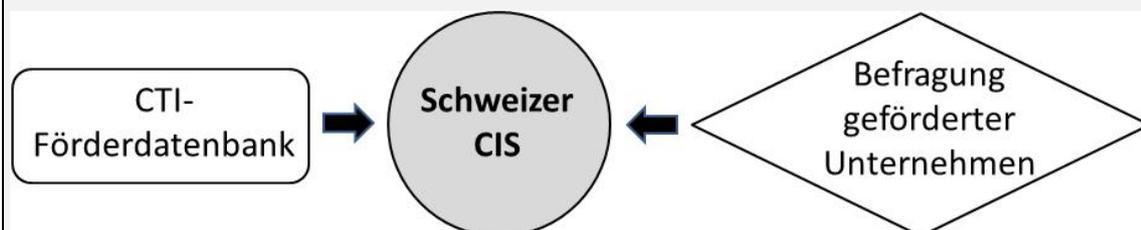
Nach den Studienautoren ist damit eine notwendige Voraussetzung für den Einsatz innovativer, ökonomischer Methoden gegeben: „Erst ein solch umfassendes Informationskonstrukt, welches das Unternehmensprofil in seiner wettbewerblichen Umgebung sowie relevante Größen der Innovationstätigkeit erfasst, macht es möglich, Zusammenhänge zwischen Innovationsanreizen, Innovationsverhalten, Wettbewerb, Ergebnissen des innovativen Prozesses der Unternehmen und staatlicher Förderpolitik aufzudecken.“ (Czarnitzki et al. 2003:67).

¹⁷ <http://www.zew.de/de/forschung/projekte.php3?action=detail&nr=374> [19.11.2011].

Box 22: Fallbeispiel Schweiz

Die Schweiz verfolgt aufgrund einer sehr verschiedenen ordnungspolitischer Orientierung im Vergleich zu Österreich eine deutlich unterschiedliche Innovationspolitik. Es gibt kaum direkte unternehmensbezogene Förderungen, wie sie etwa für die Einzelprojektförderung im Rahmen der Basisprogramme charakteristisch sind. Die wenigen F&E-Programme sind vorwiegend im Sinne eines bottom-up organisiert, wobei „the principle of indirect R&D Support of good projects, which are jointly proposed by a private and a public partner and as a main promotional policy, to our knowledge, is unique in Europe.“ (Arvanitis et al. 2010:63). Entscheidend ist dabei, dass die Subvention nicht an die Unternehmen sondern an den Wissenschaftspartner der Kooperation geht. Das Unternehmen muss seine eigenen Kosten selbst tragen (Prinzip der Ko-Finanzierung). Die Studie von Arvanitis et al. (2010) hat die Evaluierung dieses Programms zum Forschungsgegenstand. Dabei werden drei Datenquellen genutzt, um eine Matched-Pairs-Analyse zu rechnen. Die wichtigste und zentrale Quelle bildet dabei die schweizerische Innovationserhebung, welche in Österreich etwa dem CIS entspricht. Aus dieser wurden insbesondere die Unternehmen für die Kontrollgruppe ausgewählt. Weiterhin wurde die Förderdatenbank der CTI, der programmadministrierenden Agentur, verwendet. Um auch für die geförderten Unternehmen die gleichen Variablen wie für die Kontrollgruppe zur Verfügung zu haben, wurden diese zusätzlich noch mit einem Fragebogen befragt, der im Wesentlichen eine verkürzte Form des CIS darstellte. Abbildung 40 zeigt die entsprechende Datenstruktur bestehend aus drei Datenquellen.

Abbildung 40: Datenstruktur bei Arvanitis et al. (2010)



Eigene Darstellung

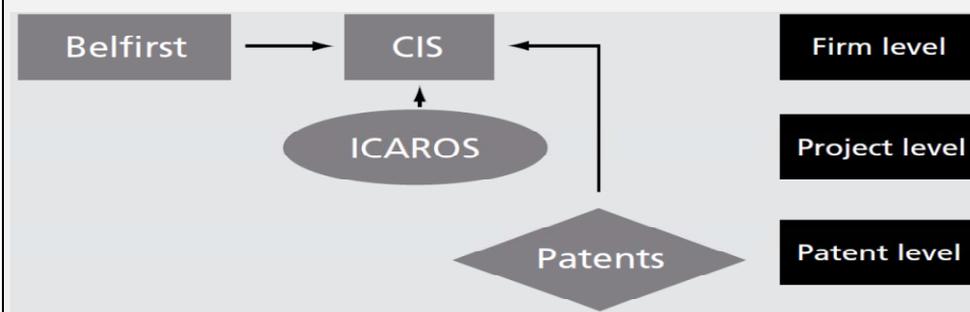
Durch die Verwendung des CIS war es möglich, folgende sechs Erfolgsgrößen hinsichtlich ihres Unterschiedes zwischen Teilnehmer- und Kontrollgruppe zu unterscheiden und für die resultierenden Unterschiede auf kausale Programmeffekte zu schließen:

1. Importance of introduced innovations from a technical point of view
2. Importance of introduced innovations from an economic point of view
3. Percentage reduction of average variable production costs due to process innovation
4. Sales of significantly improved or modified (already existing) products as percentage of total sales
5. Sales of products new to the firm or to market as percentage of total sales
6. Sales of products new to the market worldwide as a percentage of total sales

Box 23: Fallbeispiel Flandern

Aerts und Czarnitzki (2006) evaluieren in ihrer Studie die Effektivität der F&E-Subventionen, welche durch die flämische Agentur iwt vergeben werden. Wie in der deutschen und schweizerischen Studie wird auch hier eine Kontrollgruppe für einen Matching-Ansatz gebildet. Auch hier bildeten die Daten des CIS das wichtigste Datenelement, welches mit weiteren Datenquellen verknüpft wurde. Das war zunächst einmal die Belfirst-Datenbank, welche Bilanzdaten auf Unternehmensebene bereitstellt. Zweitens wurden Patentdaten des Europäischen Patentamtes mit den CIS-Daten auf Unternehmensebene zusammengefügt. Die ICAROS-Datenbank beinhaltet Informationen zum Volumen an F&E-Subventionen und Anzahl geförderter Projekte pro Jahr auf Unternehmensebene. Wie auch bei der deutschen Studie, kam es zu keiner Erhebung zusätzliche Primärdaten. Das endgültige Sample bildeten 776 Unternehmen, wobei 180 davon F&E-Subventionen von iwt erhalten hatten.

Abbildung 41: Datenstruktur bei Aerts und Czarnitzki (2006)

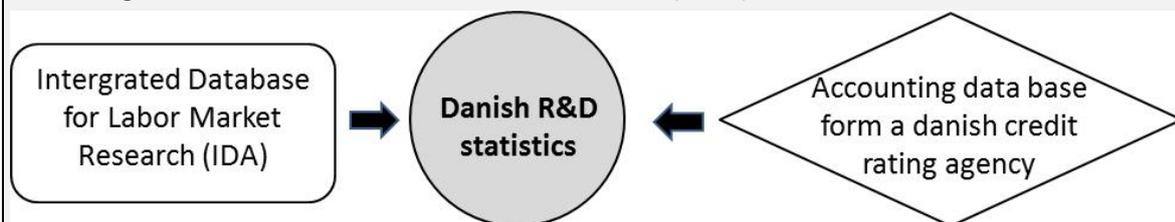


Aerts, Czarnitzki 2006

Box 24: Fallbeispiel Dänemark

Marino und Parrotta (2010) evaluieren das dänische F&E-Fördersystem unter Anwendung eines Matching-Verfahrens. Ihre Datenbasis besteht aus drei Datensets, welche alle von der Statistik Dänemark zur Verfügung gestellt wurden. Abbildung 42 zeigt die Datenstruktur der dänischen Studie. Im Gegensatz zu den anderen Fallstudien wurden hier Daten der F&E-Erhebung verwendet sowie eine Datenbasis zu Beschäftigungsdaten. Wie in den anderen Evaluationen verwendeten auch Marino und Parrotta (2010) eine Bilanzdatenbank. Durch Verknüpfung der Daten auf Mikrodatenebene entstand schließlich der Datensatz für die ökonomische Analyse.

Abbildung 42: Datenstruktur bei Marino und Parrotta (2010)



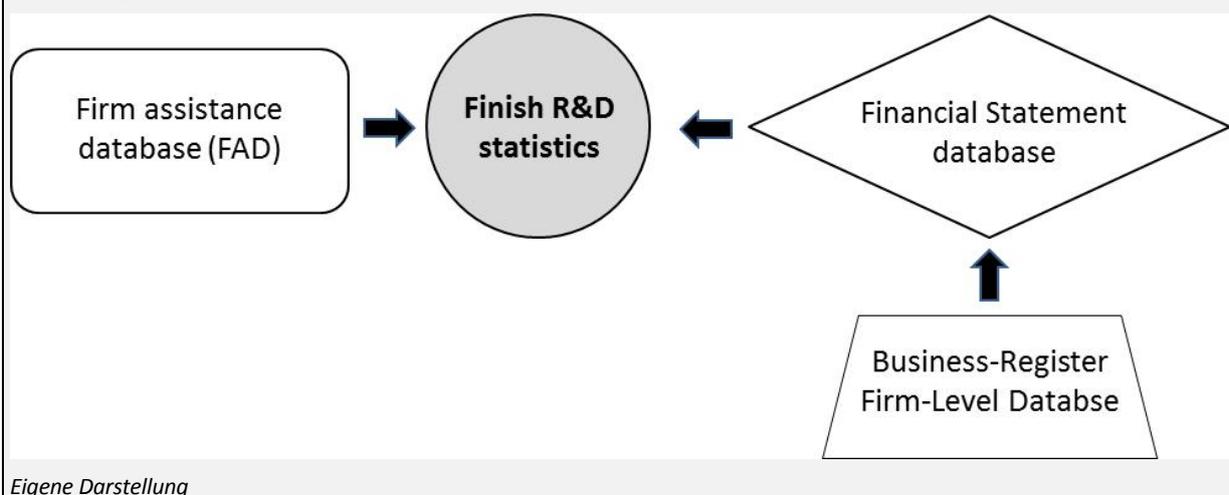
Eigene Darstellung

Box 25: Fallbeispiel Finnland

Einiö (2009) evaluiert die Effektivität der finnischen Innovationsförderung durch die Agentur Tekes. Wie im Beispiel Dänemarks wurde auch hier nicht der CIS sondern die F&E-Erhebung verwendet. Ein besonderer Vorzug der finnischen Datenlage besteht darin, dass Tekes „the only authority allocating R&D subsidies“ (Einiö 2009:6) ist und damit die „Firm assistance database“ tatsächlich alle staatlichen F&E-Förderungen enthält, die ein Unternehmen bekommen hat. Durch die Zersplitterung des österreichischen Förderwesens auf mehrere Agenturen und föderale Ebenen, ist das Problem von Messfehlern von hoher Bedeutung. Im schlechtesten Fall kommt es dazu, dass Fördereffekte, die eigentlich durch ein anderes Programm verursacht wurden dem zu evaluierenden Programm zugerechnet werden, weil keine vollständige Erfassung der Förderungen auf Unternehmensebene vorliegt. Die Ausweitung der Idee zum Aufbau einer Transparenzdatenbank von Haushalten auch für Unternehmensförderungen liegt entsprechend nahe (Rechnungshof 2011).

Als dritte Datenquelle wurden auch in dieser Studie Bilanzdaten aus zwei verschiedenen Quellen benützt. Durch Matching entstand das Daten-Set für die ökonometrische Analyse auf Basis von Mikrodaten.

Abbildung 43: Datenstruktur bei Einiö (2009)



Eigene Darstellung

Zusammenfassend lassen sich folgende Punkte festhalten: Erstens ist die Datenlage in den meisten anderen, vergleichbaren Staaten der EU wesentlich besser als in Österreich. Deshalb können dort moderne quantitative Methoden angewandt werden, welche eine wissenschaftlich nachvollziehbare Analyse hinsichtlich der der Effektivität staatlicher F&E-Förderungen leisten können. Zweitens sind Mikrodaten zur Anwendung dieser Methoden notwendig. Drittens reichen aber auch nicht die Daten z.B. nur des CIS. Vielmehr sind alle der hier vorgestellten Datensätze durch eine Verknüpfung unterschiedlicher Datenquellen auf Ebene einzelner Unternehmen gekennzeichnet. Regelmäßig werden etwa neben CIS- und F&E-erhebung auch Daten über Bilanzen und Patente hinzugespielt. Ebenfalls interessant ist die Berücksichtigung von Daten zur Beschäftigung, ist doch letztlich die Schaffung von qualitativ hochwertigen Arbeitsplätzen eines der Hauptziele jeder Innovationspolitik, jedenfalls auf makroökonomischer Ebene.

7.4. PLÄDOYERS FÜR DIE VERWENDUNG AMTLICHER UND NICHT-AMTLICHER SEKUNDÄRSTATISTIKEN

Abschnitt 7.3 hat anhand einiger Fallbeispiele gezeigt, dass die österreichische Datensituation gegenüber internationalen, westeuropäischen Standards signifikant zurückfällt. Welche Möglichkeiten gibt es, um eine Verbesserung der sich als unbefriedigend darstellenden Datenlage in Österreich, zu erreichen? Generell lassen sich zwei Ansätze unterscheiden. Erstens könnte im Rahmen von FTI-Evaluierungen eine Primärerhebung durchgeführt werden. Diese hätte neben den geförderten Unternehmen insbesondere auch nicht-geförderte Unternehmen zu befragen. Zweitens könnte, so wie auch in den internationalen Fallbeispielen in Abschnitt 7.3 dargestellt, ein Rückgriff auf bestehende Datensets erfolgen. Im Folgenden soll kurz auf die Vor- und Nachteile dieser beiden Varianten eingegangen werden.

Der Vorteil von Primärerhebungen liegt zweifelsohne in der Möglichkeit, die Fragen passgenau für die jeweilige Evaluation anzupassen. Ein Beispiel hierfür ist die Befragung in der Studie von Polt und Pointner (2005). Hier wurden einerseits 197 Programmteilnehmer des FlexCIM-Programms befragt sowie andererseits, aufbauend auf der mittlerweile nicht mehr existierenden Betriebs- und Produktdatenbank ARCS, 3786 Kontrollgruppenunternehmen. Von diesen antworteten 333 Unternehmen, wobei davon wiederum 217 Unternehmen in der Analyse berücksichtigt wurden (5,7% der Grundgesamtheit). Neben diesem Vorteil ist jedoch eine Reihe von Nachteilen zu nennen, die mehr oder weniger die Vorteile für den zweiten Ansatz zur Verbesserung der Datenlage bilden, die Verwendung von Sekundärstatistiken. Als erster Nachteil sind die (sozialen) Kosten zu nennen, die eine derartige Befragung verursacht. Neben den Kosten die auf Seiten des Evaluators anfallen, sind auch die Bürokratiekosten der Unternehmen zu berücksichtigen. Da Institutionen wie die Wirtschaftskammer bereits jetzt gegen zusätzliche Unternehmensbefragungen zahlreiche Einwände vorbringen, scheiden umfangreiche Primärbefragungen als Basis für Evaluierungen eher aus. Es sei denn, diese werden nur für einige wenige große Evaluierungen genützt. Dann bleibt es aber dabei, dass alle 5-10 Jahre einmal eine Evaluation mithilfe ökonomischer Methoden durchgeführt wird. Weiterhin ist anzumerken, dass durch den unmittelbaren Konnex zwischen Förderung und Befragung strategisches Antwortverhalten angenommen werden muss. Last but not least stellt sich auch das Problem, dass niedrige Rücklaufquoten die Ergebnisse in ihrer Aussagekraft einschränken.

Die Vor- und Nachteile von Sekundärdaten unter besonderer Berücksichtigung amtlicher Datenquellen, stellen sich wie folgt dar

- Die öffentliche Hand finanziert den Aufbau von hochqualitativen und umfassenden Datensätzen zum Innovationgeschehen in Österreich (CIS, F&E-Erhebung). Der Return dieses höchst sinnvollen öffentlichen Investments könnte durch die Verwendung dieser Daten zur Evaluation von F&E-Förderinstrumenten signifikant erhöht werden. Die Daten werden dabei rein für Zwecke verwendet, die im ureigenen Interesse von Politik und Verwaltung sind und ganz dem Geiste der Wirkungsorientierten Haushaltsführung entsprechen.

- Ein wichtiger Vorteil ist, dass Sekundärdaten per definitionem bereits existieren. Damit sind die Grenzkosten einer Verwendung für weitere Zwecke sehr niedrig. Das verbilligt die Datenbeschaffung für evaluatorische Zwecke signifikant. Freilich sind bei der Verwendung nicht-amtlicher Daten (e.g. private Unternehmensdatenbanken) entsprechende Entgelte zu entrichten. Wird jedoch eine solche Datenbank für unterschiedliche Projekte genutzt, so wird auch diese Verwendung mit jedem weiteren Projekt immer wirtschaftlicher (fallende durchschnittliche Fixkosten).
- Neben diesem Kosteneffekt ist zu bedenken, dass Unternehmen keine weiteren Fragebögen mehr beantworten müssen. Damit dient die Verwendung von Sekundärdaten der so genannten Verwaltungsvereinfachung.
- Amtliche Sekundärstatistiken im F&E-Bereich, d.h. konkret die Daten des Community Innovation Survey und die der F&E-Erhebung, verfügen über eine kontinuierlich hohe Datenqualität, wie sie durch projektbezogene Primärerhebungen niemals erreicht werden können.
- Es können höhere Fallzahlen (insbesondere bei der Kontrollgruppe) realisiert werden, womit die Ergebnisse exakter werden.
- Bei der Beantwortung der Fragen amtlicher Statistikerhebungen ist sich das Unternehmen nicht bewusst, dass diese möglicherweise auch zur Evaluation eines Programms verwendet werden, für das es Förderung erhalten hat. Damit ist strategisches Antwortverhalten unwahrscheinlich.
- Die Verwendung der amtlichen Statistik im Sinne von CIS-Daten oder Daten aus der F&E-Erhebung erhöht den kumulativen Wissensaufbau über die Wirkungsweise von F&E-Instrumenten sowohl in nationaler als auch in internationaler Perspektive. Gonzalez und Pazo (2008:374) weisen in diesem Zusammenhang darauf hin, dass sich durch die zunehmend dominante Datenkonstruktion über das Matching von Mikrodaten die Vergleichbarkeit nationaler Fallstudien deutlich verbessert hat: „These similarities may facilitate the comparability of the results, especially if the R&D policies analyzed have similar characteristics, but the use of data with different features may complicate it.“ Wie oben gezeigt, konvergiert auch die verwendete Datenbasis in Form des CIS zwischen den nationalen Evaluierungen. Eine beispielhafte Arbeit in diesem Zusammenhang ist Aerts und Schmidt (2008), die einen Vergleich von Additionalitätseffekten von F&E-Subventionen für Deutschland und Flandern im Vergleich vornahmen. Diese eignen sich nicht zuletzt deshalb besonders gut für Evaluierungszwecke, weil auch Output-Indikatoren von Innovationsprozessen erfasst werden. Im Gegensatz dazu stellt die F&E-Erhebung Daten zur Abschätzung von Input-Additionalitäten bereit. Die wiederholte Verwendung gleicher Datensätze schafft eine zunehmend bessere Kenntnis der Daten auf Seiten der Evaluatoren, so dass auch hier mit economies of scale zu rechnen ist.
- Eng damit zusammenhängend ist auch festzustellen, dass durch die Verwendung von amtlichen Daten die Transparenz und Glaubwürdigkeit von Evaluationsergebnissen zunimmt.

- Mit der Verwendung von amtlichen und nicht-amtlichen Sekundärdaten kann die österreichische Evaluierung von FTI-Maßnahmen internationalen Standards bei der Anwendung quantitativer, analytischer Methoden entsprechen und damit das Wissen über die Effektivität öffentlicher Mittelverwendung entscheidend erhöhen.

Als Nachteil von Sekundärdaten kann hier der wesentliche Vorteil von Primärerhebungen genannt werden: Es ist nicht möglich, zusätzliche, spezifische Fragen zu stellen. Allerdings muss dieses Argument auch wieder relativiert werden, weil viele wesentliche Variablen ohnehin im CIS oder in der F&E-erhebung berücksichtigt werden.

Zusammengenommen sind damit die Vorteile einer Verwendung von Sekundärdaten aus amtlichen und nicht-amtlichen Quellen weitaus größer als die Vorteile aus zusätzlichen Primärerhebungen. Wie die Ausführungen über die internationalen Fallbeispiele gezeigt haben, stellen die Daten des CIS in der Regel das grundlegende Datengerüst bereit. Allerdings sind auch die Daten der F&E-Erhebung von hoher Bedeutung.

Eine Verwendung dieser amtlichen Daten für evaluatorische Zwecke im Sinne quantitativer, analytischer Methoden sollte dabei zwei Bedingungen erfüllen: Erstens werden die Daten auf Mikroebene, d.h. auf Unternehmensebene benötigt. Zweitens sollte eine Verknüpfung mit anderen Datenquellen möglich sein. Ohne eine Verknüpfung etwa mit FFG-Förderdaten haben CIS-Daten keine sinnvolle Funktion für Evaluierungszwecke. Aber auch Verknüpfungen mit Patent-Publikations- und Unternehmensdatenbanken sind von Bedeutung. All dies ist in anderen Staaten der EU möglich. Bedenkt man die weit überdurchschnittlichen öffentlichen Aufwendungen in Österreich für F&E-Förderung, so wäre es auch hier geboten, die Evaluation dieses hohen Mitteleinsatzes durch entsprechende, bereits existierende Datenquellen zu unterstützen.

Die bisherige Situation über die Datenverfügbarkeit hinsichtlich CIS-Daten und F&E-Erhebungsdaten ist die, dass Ministerien der Statistik Austria einen Auftrag erteilen müssen, damit diese, unter streng geregelten Bedingungen, Mikrodaten zur Verfügung stellen kann. Eine Verknüpfung der amtlichen Datenquellen mit anderen Datenquellen hat in Österreich im Gegensatz zu anderen westeuropäischen Staaten der EU jedoch noch nie stattgefunden. Auch fand der eben beschriebene Vorgang einer Anweisung der Ministerien an die Statistik Austria noch nie im Zusammenhang mit Evaluierungen statt. Vielmehr handelte es sich dabei um jeweils isolierte Auswertungen der CIS-Daten z.B. für ein OECD-Projekt durch JOANNEUM RESEARCH (OECD 2009). Wie Box 26 zeigt, hat der Rat für Forschung und Technologieentwicklung in seiner Strategie 2020 eine Initiative zur Verbesserung dieser Situation gestartet. Eine weitere, damit komplementäre Forderung hat der Rechnungshof in seinem aktualisierten Positionspapier zur Verwaltungsreform erhoben. Einer der insgesamt 599 (!) Vorschläge betrifft auch eine für FTI-Evaluierungen relevante Forderung nach verbesserter Datenverfügbarkeit (siehe Box 26).

Box 26: Verbesserung der Datenlage: Dateninitiativen des RFTE und des Rechnungshofs

Die Dateninitiative des RFTE, Strategie 2010

Strategische Leitlinie 7: Datenbasis und Datenanalyse verbessern

Ein wichtiger methodischer Aspekt ist die Verbesserung der Informationsbasis für die FTI-Politik (Statistiken, Indikatoren, Mikrodaten). FTI-bezogene Daten sind teilweise nicht in den erforderlichen Abgrenzungen und Definitionen verfügbar, oder selbst bei Verfügbarkeit nicht zugänglich [Herv. d. Verf.]. (...)

Empfehlung 7.1

Gemeinsam mit den datenerfassenden Stellen (Statistik Austria, Förderagenturen, Ressorts) soll ein Prozess entwickelt werden, in dem die Erfassung und Auswertung strategisch relevanter Daten für FTI-politische Untersuchungen regelmäßig abläuft. → 2013

Monitoring: Der Rat wird die Monitoringfunktion übernehmen und gemeinsam mit den angesprochenen Akteuren eine koordinierte Lösung entwickeln.

Forderung des Rechnungshofs nach einer „Förderungs-Leistungs-Datenbank“

Auch die Initiative des Rechnungshofes im aktuellen Positionspapier zur Verwaltungsreform ist von hoher Bedeutung für FTI-Evaluierungen. Es handelt sich dabei um die Forderung den Aufbau einer zentralen „Förderungs-Leistungs-Datenbank für die Verwaltung zur konkreten Abstimmung der einzelnen Förderungsmaßnahmen und zum Vergleich des Verwaltungsaufwands der Förderungsstellen.“ Damit könne die in ökonomischen Evaluierungen verwendete Fördervariable wesentlich genauer erfasst werden. Ein Vorbild könnte hier etwa Finnland sein, wo alle F&E-Förderungen in einer Datenbank erfasst werden.

RFTE, (2010: 43).; Rechnungshof (2011:76).

In einer aktuell erschienen Studie der OECD (2011:149) fordert diese darüber hinaus auch eine Veröffentlichung aller bei der Evaluierung verwendeten Datenquellen: „Placing evaluation data in the public realm, possibly in anonymised form, could facilitate (cost-free) academic analysis and serve as a form of evaluation quality control. Furthermore, in the pluralistic policy systems that characterise OECD countries, evaluation should be expected to serve the knowledge needs of a wider set of actors beyond programme managers and public policy officials.“ Freilich scheint dies in Österreich als geradezu unerreichbares Ziel. Zunächst einmal geht es darum, dass zumindest die Evaluatoren selbst Zugang zu den Daten bekommen. Trotzdem sollte im Sinne der Transparenz und wissenschaftlichen Nachvollziehbarkeit die Forderung der OECD als Notwendigkeit zur weiteren Professionalisierung von Evaluation betrachtet werden.

Diese Studie kommt aufgrund des hier diskutierten Befunds zu dem Ergebnis, dass die Anliegen des Rats und des Rechnungshofs unumschränkt zu unterstützen sind und hier wesentliche Potenziale bislang ungenutzt bleiben. Das wohl wichtigste Argument ist, dass der Einsatz öffentlich finanzierte Datenquellen für die Evaluierung öffentlicher Mittelverwendung eine Doppeldividende lukrieren kann: Erstens werden bestehende amtliche Daten bei Grenzkosten von Null für einen zusätzlichen Zweck im öffentlichen Interesse verwendet. Zweitens kann auf Basis dieser Daten das Instrumentenportfolio der F&E-Politik hinsichtlich seiner Wirksamkeit effektiv evaluiert werden wie dies internationalen Standards entspricht.

Damit wäre die fragwürdige Asymmetrie der österreichischen FTI-Politik im Sinne eines weit überdurchschnittlichen öffentlichen Mitteleinsatzes bei unterdurchschnittlicher öffentlicher Datenverfügbarkeit zur Evaluation desselben, aufgelöst. Ob dieses Ziel über eine Änderung von datenschutzrechtlichen Bestimmungen oder über konkrete Anweisungen der Ministerien an die Statistik Austria im Zuge von Evaluierungen erreicht werden kann ist eine zu klärende juristische Frage. Wichtig wäre es aber, dass die Nutzung von Mikrodaten der Statistik Austria für F&E-Evaluierungen eine Routine wird, die nicht stets aufs Neue verhandelt werden muss. Damit wird Planungssicherheit bei Angeboten, Professionalität, Methodenkompetenz und kumuliertes Wissen über die Effektivität der österreichischen FTI-Politik befördert.

Die Verbesserung der Datenlage durch Zurverfügungstellung und Verknüpfung amtlicher F&E-Daten der Statistik Austria mit anderen Sekundärdaten würde damit auch in hohem Maße zur Realisierung einer zentralen Forderung der „Strategie der Bundesregierung für Forschung, Technologie und Innovation“ beitragen (2011:11): „Bei allen Maßnahmen steht die Wirkungsorientierung im Vordergrund. Initiativen und Programme, die nicht zu den gewünschten Wirkungen führen, werden eingestellt oder neu ausgerichtet.“

8 Resümee und Empfehlungen

In dieser Studie wurde das österreichische Evaluierungssystem im FTI-Bereich mit Schwerpunkt auf FFG- und BMVIT-Programme vor dem Hintergrund internationaler Vergleiche und nationaler Herausforderungen analysiert. Neben institutionellen Aspekten der Evaluierung wurden dabei insbesondere auch die methodischen Aspekte in den Blick genommen.

Aus der Analyse vorliegender Evaluierungen sowie aus Interviews mit Ministerien, Fördereinrichtungen und EvaluierungspraktikerInnen können folgende Schlüsse gezogen werden: Die österreichischen Evaluierungen im FTI-Bereich sind generell von hoher Qualität. Es finden darüber hinaus auch von der Anzahl her betrachtet sehr viele Evaluierungen statt. Trotzdem wurden in den Interviews auch einige Mängel angemerkt; diese betreffen unter anderem einen Mangel an ex-post Evaluierungen, Portfolioevaluierungen und Effizienzanalysen. Ebenfalls vorgebracht wurde, dass der Grad an Verbindlichkeit im Umgang mit Evaluierungsergebnissen noch gesteigert werden könnte. Einige dieser Urteile konnten auch im Rahmen eines internationalen Vergleichs von Evaluierungen bestätigt werden. Ein wichtiges Ergebnis dieses Vergleichs betrifft dabei die seltene Anwendung quantitativer, insbesondere ökonomischer Verfahren in Österreich relativ zu vergleichbaren OECD-Staaten. Während dieser Befund bereits in vergangenen Studien abzulesen war, hat sich diese Differenz weiter akzentuiert. Diese Trends wurden vor dem Hintergrund zweier wichtiger institutioneller Neuerungen (Themenmanagement der FFG, Wirkungsorientierte Haushaltsführung) dargestellt und hinsichtlich ihrer Bedeutung für FTI-Evaluationen diskutiert.

Zunächst zum Themenmanagement. Die bisherige Evaluierungstradition war und ist von der Evaluierung einzelner Programme geprägt (Programmevaluierung). Dies ist zunehmend unzureichend angesichts der starken Wechselwirkungen von Förderinstrumenten als auch der entwickelten Technologien und Innovationen. Darauf reagiert die Politik als auch die Fördereinrichtungen zunehmend durch einen ‚systemhaften‘ Zugang. So ist etwa durch das in naher Zukunft zu implementierende sogenannte „Themenmanagement“ der FFG eine tiefgreifende Veränderung der Evaluierungspraxis zu erwarten. Unter Themenmanagement versteht man den an einzelne Themen (z.B. Mobilität, Energie) angepassten Einsatz von standardisierten Instrumenten (z.B. Einzelprojektförderung, Kooperationsförderung) der FFG zwecks Förderung der Innovationsperformance im jeweiligen Themenbereich. Dabei steht die gesamthafte Wirkung der FFG-Instrumente auf die Entwicklung von verschiedenen, thematisch verbundenen Branchen und unterschiedlichen Akteuren im Zentrum des Erkenntnisinteresses.

Bei der Wirkungsorientierten Haushaltsführung (WHF) handelt es sich um eine Maßnahme im Rahmen der zweiten Etappe der Haushaltsrechtsreform, welche mit 1.1.2013 in Kraft treten wird. Ebenfalls Teil der Haushaltsrechtsreform ist die Einführung eines „Gender Budgeting“. Die WHF hat zunächst wenig mit herkömmlichen Evaluierungen im FTI-Bereich zu tun. Es handelt sich dabei um einen internen Verwaltungsprozess, d.h. die mit der WHF verbundenen Aktivitäten erfolgen durch

die Verwaltung selbst. Diese definiert in einem ersten Schritt Wirkungsziele, zugeordnete Maßnahmen zur Erreichung dieser Wirkungsziele sowie Indikatoren bzw. Kennzahlen.

Weiterhin setzt sich die jeweilige Verwaltungseinheit numerisch bestimmte Ziele (Meilensteine), welche in einem jährlich stattfindenden Ist-Soll-Vergleich auf ihren Zielerreichungsgrad hin überprüft werden sollen. Dabei sind zwei Probleme zu beachten. Erstens erfolgt die Identifikation der Wirkungen FTI-politischer Intervention im Rahmen der WHF durch einen deskriptiven Vorher-Nachher-Vergleich. Dies würde fälschlicherweise unterstellen, dass alle Veränderungen des Innovationssystems im Sinne der definierten Indikatoren direkt und ausschließlich auf die jeweilige politische Intervention zurückzuführen wären. Da diese Annahme offensichtlich fragwürdig ist, steht man vor dem Problem herauszufinden, welcher Anteil der Entwicklung durch die Innovationspolitik verursacht worden ist. Zweitens sollte darauf geachtet werden, dass die Zielsetzungen der WHF in inhaltlicher und zeitlicher Hinsicht nicht im Widerspruch zur Forderung nach vermehrter Förderung risikoreicher F&E Projekte stehen. Werden vermehrt radikale Innovationen gefördert, so ist auch mit einer erhöhten Volatilität von Output-Variablen der Förderung zu rechnen. Dies würde die Wahrscheinlichkeit zur Verfehlung der Ziele zwar nicht im mehrjährigen Durchschnitt, aber auf jährlicher Basis erhöhen

Im Methodenteil präsentiert die Studie wichtige neuere Entwicklungen im Bereich qualitativer und quantitativer Methoden, die auch für den Einsatz bei FTI-Evaluationen geeignet sind. Innovative Fokusgruppendesigns (quanti-quali Designs, partizipative Designs, etc.) werden vorgestellt, deren Stärke darin liegt, die Probleme traditioneller Fokusgruppenforschung (group conformity, preference toward agreement, etc.) zu reduzieren.

Qualitative soziale Netzwerkanalysen oder die Kombination aus quantitativen und qualitativen Elementen der Netzwerkforschung sind insbesondere vor dem Hintergrund von Wirkungsanalysen FTI-politischer Netzwerkiniciativen vielversprechende Verfahren. Sie ermöglichen die Erforschung von Erfolgsfaktoren von Innovationsnetzwerken, Netzwerkpraktiken und Netzwerkinterpretationen – wichtige Aspekte, die von der quantitativen Netzwerkanalyse nur eingeschränkt bearbeitet werden können.

Bei den quantitativen Verfahren wurde der mikroökonomische Ansatz ausführlich erörtert. Dessen Bedeutung liegt vor allem darin, dass hier eine Kernfrage jeder Evaluation explizit gemacht und im Zentrum des Erkenntnisinteresses steht: Wie hätten sich die geförderten Unternehmen, ForscherInnen, etc. entwickelt, hätten diese keine Förderung bekommen? Dies ist eine kontrafaktische Frage nach der kausalen Wirkung politischer Intervention. Eine Evaluierung, die keine Antwort auf diese Frage gibt, verfehlt ein zentrales Anliegen von Evaluation. Das Problem ist jedoch, dass die kontrafaktische Situation und damit die Antwort auf diese Frage nicht beobachtet werden kann. Deshalb ist es auch falsch, wenn von der Messung kausaler Effekte gesprochen wird; diese können nur über das Treffen bestimmter, nicht überprüfbarer Annahmen, geschätzt werden. Diese Einsicht führt zur Erkenntnis, dass weder deskriptive und buchhalterische Ansätze noch einfaches Nachfragen bei den geförderten Subjekten dieser komplexen Fragestellung gerecht werden. Es braucht den Einsatz komplexerer mikroökonomischer Methoden, wobei mit dem

Matching-Ansatz und dem Kontrollfunktionsansatz die zwei am häufigsten angewandten Methoden in diesem Zusammenhang dargestellt werden.

Die steigende Bedeutung der Innovationspolitik in den OECD-Staaten zeigt sich nicht zuletzt darin, dass die damit verfolgten Ziele zunehmend komplexer werden. Mehrdimensionale sozioökonomische Wirkungsdimensionen von innovationspolitischen Maßnahmen gewinnen an Bedeutung relativ zu eindimensionalen ökonomischen Zielsetzungen. Dies stellt neue Herausforderungen an die Evaluierungspraxis. Welche Ziele können bei nüchterner Betrachtung der Interventionkapazitäten des politischen Systems formuliert werden? Wie sollen diese Ziele gemessen und deren Erreichung evaluiert werden? Der Einsatz von kombinierten Methoden sowie interdisziplinären und partizipativen Verfahren ist momentan ein vielversprechender Weg, um der Komplexität und Kontextbezogenheit sozialer und gesellschaftspolitischer Zielsetzungen gerecht zu werden. Zweifelsohne sind die erwähnten Fragestellungen aber noch keineswegs abschließend beantwortet worden.

Als letzten Punkt erörtert diese Studie ausführlich die entscheidende Frage der Datenverfügbarkeit für evaluatorische Zwecke. Dabei wird zunächst herausgestellt, dass Methoden und Daten in einem unauflöselichem Zusammenhang zueinander stehen: Es nützen die besten Methoden nichts ohne Daten! In Österreich besteht jedoch leider eine Situation unbefriedigender Datenverfügbarkeit für eine Reihe von wichtigen evaluatorischen Fragestellungen. Insbesondere sind Mikrodaten der wichtigen amtlichen F&E-Statistikwerke (Community Innovation Survey, F&E-Erhebung) für evaluatorische Zwecke nicht verfügbar. Ein internationaler Vergleich macht deutlich, dass dies in vielen vergleichbaren OECD-Staaten, wie z.B. Deutschland, Schweiz oder Finnland hingegen der Fall ist. Die Folge ist ein besserer Informationsstand über die Effektivität staatlicher F&E-Förderung durch die Anwendung quantitativer, analytischer Verfahren. Hier hat Österreich eindeutig Nachholbedarf. Die einschlägigen Kosten sind dabei sehr niedrig weil diese Daten bereits bestehen. Demgegenüber ist der Ertrag potenziell sehr hoch, weil die Identifikation ineffektiver Förderprogramme einen signifikanten Beitrag zur Steigerung der Effizienz des öffentlichen Sektors zu leisten vermag.

Zusammenfassend ergeben sich folgende Handlungsempfehlungen:

Weiterentwicklung der Evaluationsmethoden zur besseren Wirkungsmessung. Für die österreichische Situation bedeutet dies den vermehrten Einsatz von avancierteren quantitativen Verfahren, die stärker als die bisher angewandten auf Wirkungsmessung abzielen, sowie von neueren qualitativen Methoden. Hier ist ein ausgewogener Methodenmix von qualitativen und quantitativen Methoden anzustreben.

EvaluatorInnen sollten durch die Ausschreibungsbedingungen und die evaluativen Fragestellungen ermutigt werden, innovative neue methodische Ansätze auszuprobieren und zur kontinuierlichen Weiterentwicklung bestehender Methoden beizutragen.

- **Verbesserung der Datenlage für Evaluierungszwecke.** Die Verbesserung der Datenlage stellt eine wesentliche Vorbedingung für wirkungsorientierte Evaluierung von FTI-Politik dar. Die Situation bezüglich Datenlage ist nicht nur in anderen Ländern (z.B. DEU, FIN, CH), sondern auch in anderen Politikfeldern (Arbeitsmarkt-, Exportstatistiken) in Österreich deutlich besser. Daten, die mit öffentlichen Mitteln erhoben wurden sollten für die Kontrolle der staatlichen Mittelverwendung eingesetzt werden können (wie der Community Innovation Survey oder die F&E-Erhebung) und in Form anonymisierter Mikrodaten zugänglich gemacht werden. Die Möglichkeit zur Verknüpfung mit weiteren Sekundärdatenquellen ist ebenfalls eine notwendige Voraussetzung für die Anwendung von state-of-the-art Methoden der Evaluationsforschung. Gegebenenfalls sind Änderungen der rechtlichen Grundlagen des Datenzugangs in Österreich anzustreben.
- **Verbesserte Abstimmung der Programme und Instrumente durch vermehrte Portfolioevaluierung.** Das österreichische Innovationsförderungssystem ist durch eine Vielzahl an Akteuren und Programmen gekennzeichnet. Doppelförderungen und Förderkonkurrenz sind mögliche unerwünschte Nebenwirkungen eines solchen Fördersystems. Umso wichtiger wird aber die Berücksichtigung der Einbettung des zu evaluierenden Programms in die bestehende Förderlandschaft, d.h. eine Portfolioevaluierung. Diese sollte vermehrt Berücksichtigung finden, um den Instrumentenmix zumindest in einem iterativen Prozess aufeinander abzustimmen.
- **Verbesserung der Qualität und des Nutzens von Evaluierungen durch konzise und akkordierte Formulierung der „Terms of References“ (ToR).** Die in den Ausschreibungen von Evaluierungen formulierten Anforderungen und evaluatorischen Fragestellungen (Terms of References) haben einen zentralen Einfluss auf die Qualität von Evaluierungen. Die Studie kommt zu dem Ergebnis, dass die ToR noch weiter verbessert werden können. Drei Maßnahmen hierzu sind: (1) die Einbeziehung der Förderagenturen in den Prozess der Formulierung der ToR, (2) die klare Arbeitsteilung zwischen Festlegung der ToR durch die Auftraggeber von Evaluierung und die Entscheidung über die Methode durch die EvaluatorInnen sowie (3) realistische Ansprüche an den Umfang der zu beantwortenden Fragen.
- **Klare Zielformulierung in der Programmkonzeption als Voraussetzung für Wirkungsabschätzung.** Wirkungsanalysen brauchen eindeutig definierte und messbare Ziele, die den EvaluatorInnen Aussagen in Bezug auf relevante Veränderungen und Wirkungen ermöglichen. Werden Ziele zu abstrakt formuliert (z.B. Steigerung der Wettbewerbsfähigkeit) erschwert bzw. verunmöglicht dies die Wirkungsabschätzung.

- **Verstärkte Reflexion von sozioökonomischen Indikatoren.** Um die Auswirkungen von Programmen umfassend zu analysieren, sollten auch Wirkungstypen die über ökonomische Effekte hinausgehen (z.B. ökologische Nachhaltigkeit, Nutzerfreundlichkeit, Gender Equality, Public Awareness) in die Wirkungsanalyse einbezogen werden. Hierfür gibt es jedoch noch wenige bis keine allgemein akzeptierten Indikatoren – diese müssen sowohl in Bezug auf die unmittelbaren Ziele sowie den Inhalt und Kontext des Programms formuliert und reflektiert werden. Sollten solche Ziele in Programmen vorkommen bzw. der Evaluierung gewünscht sein, müssten sie dafür im Kontext des jeweiligen Programmes geeignet operationalisiert werden. In der Entwicklung neuer Metriken für die sozio-ökonomischen Indikatoren liegt ebenfalls eine Herausforderung für die Zukunft.
- **Mehr lernen von Evaluierungen durch breite und öffentliche Diskussion.** Um den Einfluss von Evaluierungen zu erhöhen, ist eine breite und möglichst zeitnahe Veröffentlichung von Evaluierungen wünschenswert.

Referenzen

- Aerts, K.; Schmidt, T. (2008): Two for the price of one? Additionality effects of R&D subsidies: comparison between Flanders and Germany. In: *Research Policy*, 37, 806-822.
- Aerts, K.; Czarnitzki, D. (2004): The impact of public R&D-funding in Flanders. *Iwt studies*, 54.
- Aiginger, K. (2009): Evaluation of government funding in RTDI from a systems perspective in Austria. Synthesis Report.
- Angrist, J.; Pischke, J.E. (2010): The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. In: *Journal of Economic Perspectives*, 2, 3-10.
- Angrist, J.; Pischke, J.E. (2009): *Mostly harmless econometrics. An empiricist's companion.* Princeton.
- Arnold, E. (2005): Evaluating research and innovation policy: a systems world needs systems evaluation. In: *Research Evaluation*, 1, 3-17.
- Arvanitis, A. (2010): Impact of Swiss technology policy on firm innovation performance: an evaluation based on matching approach. *Science and Public Policy*, 1, 63-78.
- Badinger, H. (2011): A new growth strategy for Europe? *Volkswirtschaftliche Tagung ONB 2011.* Online:
http://www.oenb.at/de/presse_pub/period_pub/volkswirtschaft/vowitagung/volkswirtschaftliche_tagung_2011.jsp#tcm:14-239363, 1.12.2011.
- Bauer, T.; Grave, B. (2011): Performance related funding of universities – does more competition lead to grade inflation? *RUHR Economic Papers*, 288.
- Bauer, T. et al. (2009): *Empirische Wirtschaftsforschung. Eine Einführung.* Berlin, Heidelberg.
- Berube, C.; Mohnen, P. (2007): Are firms that received subsidies more innovative? *UNI-Merit Working Paper Series*, 15.
- Biegelbauer, P. (2010): 25 Jahre Staatliche Steuerungsversuche in der österreichischen FTI-Politik: Neue Lösungen, alte Probleme. In: Biegelbauer, P. (Hrsg.): *Steuerung von Wissenschaft? Die Governance des österreichischen Innovationssystems.* Innsbruck et al., 67-108.
- Blundell, R.; Costa Dias, M. (2008): Alternative approaches to evaluation in empirical microeconomics. *CEMMAP Working Paper*, 26.
- BMF (2011): *Strategiebericht zum Bundesfinanzrahmengesetz 2012-2015.*
- Buigues, P.A., Sekkat, K. (2011): Public subsidies to business. An international comparison. In: *Journal of Industry, Competition and Trade*, 11, 1-24.
- Cameron, C.; Trivedi, P. (2005): *Microeconometrics. Methods and applications.* Cambridge.
- Canter, U.; Kösters, S. (2011): Picking the winner? Empirical evidence in the targeting of R&D subsidies to start ups. In: *Small Business Economics.* Online first.
- CIA (2010), *The World Factbook*, Washington: CIA. Online: www.cia.gov/library/publications/theworld-factbook/index.html.
- Cincera, M. et al. (2009): Efficiency of public spending in support of R&D activities. *Economic Papers*, 376.
- Clarysse, B.; Knockaert, M. (2009): Fishing for carps in a goldfish pond? An analysis of R&D subsidy applicants and beneficiaries. *Iwt*, 64.

- Claussen, T. (2010): Do subsidies have positive impacts on R&D and innovation activities at the firm level? TIK Working Papers on Innovation Studies, 0615.
- Cornet, M. et al. (2006): Do innovation vouchers help SMEs to cross the bridge towards science? CPB Discussion Paper, 58.
- Cornet, M.; Webbink, H. (2004): Towards evidence based policy. CPB Document, 48b.
- Cozzens, S. et al. (2002): Outcomes of innovation. Evaluating the distributional consequences of science and technology policies and programs. In: Research Evaluation, 2, 101-107.
- Cozzens, S. (1999): Are new accountability rules bad for science? In: Issues in Science and Technology.
- Czarnitzki, D. (2010): Evaluation of public R&D policies: A cross-country comparison. ZEW Discussion Paper, 10.
- Czarnitzki, D. et al. (2003): Einsatzmöglichkeiten neuer quantitativer Ansätze zur Wirkungsanalyse öffentlicher Forschungsförderung – Eine Grundlagenstudie. Zentrum für Europäische Wirtschaftsforschung.
- Dasgupta, P. (2007): Economics. A very short introduction. Oxford.
- Einoö, E. (2009): The effect of government subsidies on private R&D: evidence from geographic variation in support program funding. Helsinki Centre for Economic Research, Discussion Papers, 263.
- Eckl, V.; Engel, D. (2011): How to benefit from publicly funded pre-competitive research: an empirical investigation for Germany's ICR-program. In: Journal of Technology Transfer, 36, 292-315.
- Econsearch (2008): Cost benefit analysis of a GWRDC project cluster: Vine physiology-flowering.
- Europäische Kommission (2009): Leitlinien zur Folgenabschätzung. Online: http://ec.europa.eu/governance/impact/commission_guidelines/docs/iag_2009_de.pdf, 24.10.2011
- Europäische Kommission (2009): Assessing Social Impacts. Ref. Ares(2009)326974 - 17/11/2009. Online: <http://ec.europa.eu/social/main.jsp?catId=760&langId=en&preview=cHJldmld0VtcGxQb3J0YWwh>, 24.10.2011
- European Commission: Europe 2020 Flagship Initiative Innovation Union. SEC (2010) 1161.
- Fteval (2005): Standards der Evaluierung in der Forschungs- und Technologiepolitik. Kommentierte Version.
- Fahrenkrog G. et al. (Hrsg.) (2002): RTD Evaluation Toolbox.
- Falk, R. (2006): Measuring the effects of public support schemes on firms' innovation activities. In: Research Policy, 36, 665-679.
- Falk et al. (2009): Tax incentive schemes for R&D investment. Evaluation of government funding from a systems perspective in Austria. Wien.
- Franke, K.; Wald, A. (2006): Möglichkeiten der Triangulation quantitativer und qualitativer Methoden in der Netzwerkanalyse. In: Hollstein, B.; Straus, F.(Hg.), Qualitative Netzwerkanalyse. Konzepte, Methoden, Anwendungen, Wiesbaden, 153–177. online: http://www.ssoar.info/ssoar/files/dgs/32-2004/4395_Franke_Triangulation.pdf
- Garbarino, S.; Holland, J. (2009): Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results. Issues Paper online: <http://www.gsdr.org/docs/open/EIRS4.pdf>.

- Gassler, H.; Polt, W. (2006): Alter Wein in neuen Schläuchen? Zur aktuellen Diskussion um technologiepolitische Schwerpunktsetzungen. In: Reith, R. et al. (Hrsg.): Innovationskultur in historischer und ökonomischer Perspektive. Modelle, Indikatoren und regionale Entwicklungslinien. Innsbruck et al., 159-173.
- Gault, F. (2010): Innovation Strategies for a Global Economy, Development, Implementation, Measurement and Management, Cheltenham.
- Gault, F. (2011): Social impacts of the development of science, technology and innovation indicators. UNU-MERIT Working Papers Nr. 2011-008.
- Georghiou L. (2007): What lies beneath: avoiding the risk of under-evaluation. In: Science and Public Policy, 10, 743-752.
- Georghiou L. (2004): Impact and additionality of innovation policy. *iwt studies*, 40, 57-65.
- Georghiou L, Roessner D (2000): Evaluating technology programs: tools and methods. In: Research Policy, 29, 657-678.
- Godin, B. (2002): The rise of innovation surveys. Measuring a fuzzy concept. Working Paper 16, Montreal.
- Godin, B. (2006): Statistics and Science, Technology and Innovation Policy: How to get relevant indicators. OECD Blue Sky II Conference. Ottawa. Konferenzbeitrag. Online: <http://www.oecd.org/dataoecd/48/15/37483949.pdf>, 25.10.2011
- Godin, B. und Doré, C. (2006): Measuring the Impact of Science: Beyond the Economic Dimension. Working Paper, online: http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf, 24.10.2011.
- Gonzalez, X.; Pazo, C. (2008): Do public subsidies stimulate R&D spending? In: Research Policy, 37, 371-389.
- Görg, H.; Strobel, E. (2007): The Effect of R&D Subsidies on Private R&D. In: *Economica*, 74, 215-234.
- Gottweis, H.; Latzer, M. (2006): Forschungs- und Technologiepolitik. In: Politik in Österreich. Das Handbuch. Wien, 711-725.
- Green, W (2011): *Econometric Analysis*. 7.Aufl., Boston et al.
- Grim, B. J. et al. (2006): Focused Group Interviews as an Innovative Quanti-Qualitative Methodology (QQM): Integrating Quantitative Elements into a Qualitative Methodology. The Qualitative Report, 3, 516-537. online: <http://www.nova.edu/ssss/QR/QR11-3/grim.pdf>, 25.10.2011.
- Handbook of Cost-Benefit-Analysis. Online: <http://www.finance.gov.au/obpr/cost-benefit-analysis.html>, 1.12.2011.
- Hanusch, H. (2011): *Nutzen-Kosten-Analyse*. 3.Aufl. München.
- Hauser, R. M.; Sheridan, J.; Warren, J.R. (1998): Choosing a Measure of Occupational Standing. How Useful are Composite Measures in Analyses of Gender Inequality in Occupational Attainment? CDE Working Paper 96-10. online: <http://www.ssc.wisc.edu/cde/cdewp/96-10.pdf>.
- Heckman, J. et al. (1999): The economics and econometrics of active labor market programs. In: Ashenfeter, A.; Card, D. (Hrsg.): *Handbook of labor economics*, 1865-2097.
- Heckman J., Smith J. (1995): Assessing the case for social experiments. In: *Journal of Economic Perspectives*, 9, 85-110.

- Hollstein, B. (2007): Netzwerke, Akteure und Bedeutungen. Zur Integration qualitativer und quantitativer Verfahren in der Netzwerkforschung. In: Karl-Siegbert Rehberg (Hg.), Die Natur der Gesellschaft. Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006. Frankfurt/Main.
- Hollstein, B.; Straus, F. (Hg.) (2006): Qualitative Netzwerkanalyse. Konzepte, Methoden, Anwendungen, Wiesbaden.
- Hussinger, K. (2008): R&D and subsidies at the firm level: an application of parametric and semiparametric two-step selection models. In: Journal of Applied Econometrics, 23, 729-747.
- Imbens, G., Wooldridge, J. (2009): Recent developments in the econometrics of program evaluation. In: Journal of Econometric Perspectives, 47, 5-86.
- INNO APPRAISAL (2010). Online: <http://www.proinno-europe.eu/appraisal>, 1.12.2011.
- Jaffe, A. (2002): Building programme evaluation into the design of public research-support programmes. In: Oxford Review of Economic Policy, 1, 22-34.
- Janger, J. (2010): Forschungs- und Innovationspolitik nach der Wirtschaftskrise. WIFO-Monatsberichte, 8, 675-689.
- Kahn, R. L.; Antonucci, T. C. (1980): Convoys Over the Life Course: Attachment, Roles, and Social Support. In: Baltes, P. B.; Brim, O. G. (Hg.), Life-Span Development and Behavior. New York, 383-405.
- Khandker, S.R. et al. (2010): Handbook on impact evaluation. Quantitative methods and practices. Washington.
- Klette, T. (2000): Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies. In: Research Policy, 29, 471-495.
- KMU-Forschung Austria (2011): FFG-Wirkungsmonitoring 2010.
- Kuhlmann, S. (2003): Evaluation of research and innovation policies: a discussion of trends with examples from Germany. In: International Journal of Technology Management, 2/3/4, 131-149.
- Lalonde, R. (1986): Evaluating the econometric evaluations of training programs with experimental data. American Economic Review, 4, 604-620.
- Laudel, G.; Gläser, J. (2011): Academic careers and how to find research excellence. In: Zeitschrift Plattform Forschungs- und Technologieevaluierung (fteval) 36.
- Leamer, E. (1983): Let's take the con out of econometrics. In: American Economic Review, 1, 31-43.
- Leo, H. et al. (2006): WIFO-Weißbuch. Mehr Beschäftigung auf Basis von Innovation und Qualifikation. WIFO-Weißbuch, Teilstudie 8.
- Lundvall, B.A.; Borras, S. (2005): Science, technology, and innovation policy. In: Fagerberg, J. (et al.) (Hrsg.): The Oxford handbook of innovation. Oxford, 599-631.
- Link, A., Scott, J. (2011): The theory and practice of public-sector R&D economic impact analysis. Planning Report 11-1. National Institute of Standards and Technology.
- Macilwain, C. (2010): What science is really worth. In: Nature, 10, 682-684.
- Marino, M.; Parrotta, P. (2010): Impacts of public funding to R&D: Evidence from Denmark. Paper presented at the DRUID Summer Conference 2010.
- Mayring, P. (2002): Einführung in die Qualitative Sozialforschung. Weinheim und Basel.
- Nowotny, E.; Zagler, M. (2009): Der öffentliche Sektor. Einführung in die Finanzwissenschaft. 5. Aufl. Berlin, Heidelberg.

- OECD (2011): Business innovation policies.
- OECD (2010): The OECD Innovation strategy. Getting a head start on tomorrow. Paris.
- OECD (2009): Innovation in firms. A microeconomic perspective. Paris.
- OECD (2008): Science, Technology and Industry Outlook.
- OECD (2008a): Handbook on Constructing Composite Indicators. Methodology and User Guide. Online: <http://www.oecd.org/dataoecd/37/42/42495745.pdf>.
- Pichler, R. (2009): Institutionelle Dimensionen von Evaluierung in Österreich. In: Widmer, T. et al. (Hrsg.): Evaluation. Ein systematisches Handbuch. Wiesbaden, 40-51.
- Polt, W.; Pointner, W. (2005): Diffusionsorientierte Technologiepolitik. Eine vergleichende Wirkungsanalyse für Österreich, die Schweiz, Deutschland und die USA. Graz.
- Powell, J. (2006): Benefit-Cost Analysis: Overview. In: fteval Newsletter, 31, 10-24.
- Rat für Forschung und Technologieentwicklung (2010): Strategie 2020.
- Rechnungshof (2011): Positionen. Verwaltungsreform 2011.
- Rechnungshof (2007): BMF. BMWA. Ausgewählte Maßnahmen der indirekten Forschungsförderung. Bund 2007/6.
- Reinkowiski, E. (2006): Mikroökonomische Evaluation und das Selektionsproblem. Ein anwendungsorientierter Überblick über nichtparametrische Verfahren. In: Zeitschrift für Evaluation, 2, 187-226.
- Republik Österreich (2011): Der Weg zum Innovation Leader. Strategie der Bundesregierung für Forschung, Technologie und Innovation.
- Rhomberg, W. et al. (2006): Neue Entwicklungen im Bereich der Wirkungsanalyse und – abschätzung FTI-politischer Maßnahmen. Austrian Research Centres.
- Ruegg, R.; Feller, I. (2003): A toolkit for evaluating public R&D investment. Models, methods and findings from ATP's first decade. Online: <http://www.atp.nist.gov/eao/gcr03-857/contents.htm>, 1.12.2011.
- Sachverständigenrat (2010): Jahresgutachten 2009/10.
- Schiefer, A. (2011): Forschung und experimentelle Entwicklung (F&E) im Unternehmenssektor 2009 – Teil1. In: Statistische Nachrichten, 10, 974-997.
- Schibany, A. (Koordination), M. Berger, B. Dachs, M. Dinges, H. Gassler, J. Janger, F. Kampik, A. Reinstaller, G. Streicher, S. Turecek, F. Unterlass, G. Zahradnik (2011), Österreichischer Forschungs- und Technologiebericht 2011, Wien.
- Schmidt, C. (2007): Policy evaluation and economic policy advice. Ruhr Economic Papers, 1.
- Schmidt, C. (1999): Knowing what works. The case for rigorous program evaluation. IZA Discussion Paper Series, 77.
- Schulte, K. (1997): Freundschaftsbeziehungen in persönlichen Netzwerken. In: Lang, S. (Hg.), Kulturelle Identität, soziale Netzwerke und Kognition. Berichte ethnologischer Forschung aus Köln. Wiesbaden, 11-36.
- Straus, F. (2002): Netzwerkanalyse. Gemeindepsychologische Perspektiven für Forschung und Praxis. Wiesbaden.
- Streicher, G. (2007): Additionality of FFG-funding. InTeReg Working Paper, 49.
- Takalo T. (2008): Evaluating innovation policy: a structural treatment effect model of R&D subsidies. Bank of Finland Research Papers, 7.

- Tichy, G. (2009): Was ist das „Österreichische“ an der österreichischen FTI-Politik? In: Leitner, K.H. et al. (Hrsg.): Innovationsforschung und Technologiepolitik in Österreich. Innsbruck, 255-272.
- Verbeek, M. (2008): A guide to modern econometrics. 3.Aufl. Chichester.
- Vicsek, L. (2007): A Scheme for Analyzing the Results of Focus Groups. In: International Journal of Qualitative Methods 6, 4, 20-34.
- Vicsek, L. (2010): Issues in the Analysis of Focus Groups: Generalisability, Quantifiability, Treatment of Context and Quotations. The Quantitative Report, 1, 122-141. Online: <http://www.nova.edu/ssss/QR/QR15-1/vicsek.pdf>, 24.10.2011.
- Vinnova Analysis (2007): Effects of Swedish Traffic Safety Research 1971-2004. Online: <http://www.vinnova.se/upload/EPiStorePDF/va-07-10.pdf>, 24.10.2011.
- Wolff, G.; Reinthaler, V. (2008): The effectiveness of subsidies revisited: accounting for wage and employment effects in business R&D. In: Research Policy, 37, 1403-1412.
- Zimmermann, H.; Henke, K.D. (2005): Finanzwissenschaft. München.
- Zinöcker, K. (2009): Evaluation von Forschungs- und Technologiepolitik in Österreich. In: In: Widmer, T. et al. (Hrsg.): Evaluation. Ein systematisches Handbuch. Wiesbaden, 295-304.
- Zinöcker, K. (2007): Evaluating Austria's R&D policies. Some personal comments. In: Zinöcker, K. et al. (Hrsg.): Evaluation of Austrian research and technology policies. A summary of Austrian evaluation studies from 2003-2007. Wien.
- Zuckerman-Parker, M.; Shank, G. (2008): The Town Hall Focus Group. A New Format for Qualitative Research Methods. The Qualitative Report, 4, 630-635. Online: <http://www.nova.edu/ssss/QR/QR13-4/zuckerman-parker.pdf>, 24.10.2011.

Anhang A : Indikatoren für verschiedene sozio-ökonomische Wirkungsdimensionen

Impact	Indicator	
	Indicator nominal	Indicator ordinal/numerical
Impact on culture – knowledge (the knowledge and understanding of the ideas and reality by the individuals acquired through formal or informal mechanisms)		The rate of university, technical and professional graduation in the sciences
		Academic results in the sciences
		The level of understanding of scientific concepts
Impact on culture – Know-how, Intellectual skills	The development of new skills: creativity, critique, analysis and synthesis	The level of mastery of the new acquired skills
Impact on culture – Know-how, Practical skills	The ability to identify and solve certain problems of a mechanical or technical nature at work or at home The presence of new technologies at work or at home	The frequency and duration of use of new technologies at work and at home
Impact on culture – Attitudes and Interest (For science in general, scientific insitutions, S&T controversies, scientific news and culture in general)	The participation in scientific activities	The number of hours dedicated by an individual listening to or watching scientific programming on the television/radio, to scientific leisure activities (reading, clubs etc.) The number of hours dedicated to reading newspapers and magazines on science and technology The level of coverage of science news in the media The level of acceptance and innovation of S&T (GMOs, cloning, etc.) The number of visitors to S&T museums

Impact	Indicator	
	Indicator nominal	Indicator ordinal/numerical
Impact on culture – Values and Beliefs	Values (moral, intellectual and professional) and beliefs (religious, spiritual and family)	
Impact on society- Well being and quality of life	Improving the social conditions of individuals Improving the economic conditions of individuals	Revenues of individuals
Impact on society- Social implication	Engagement within associations working on scientific questions	
Impact on society- Practices		The number of individuals having modified one or several customs or lifestyle habits (food, sexuality, activities)
Impact on society- Organisations – Speeches, interventions and actions	The appearance of new discourses on S&T The appearance of new styles of intervention or the solution to social problems	
Impact on policy- Decision maker – Alertness, interests, attitudes	A new interest or attitude towards questions of public interest involving S&T	
Impact on organisations Planning – Objectives	New strategic orientations, missions or objectives	
Impact on organisations Planning – Administrative organisation	An administrative restructuring	The number of people affected by the restructuring
Impact on organisations- Organisation of work – tasks	The allocation of staff (work division)	The degree of specialisation of the jobs
Impact on organisations- Organisation of work – Automation	Acquisition of advanced production techniques	
Impact on organisations- Organisation of work – Computing	The architecture of the computer network	The number of computer jobs in the organisation
Impact on organisations- Administration -Management		The level of qualification and the years of experience of management personnel
Impact on organisations- Human Resources – Workforce		The number of new employees in R&D
Impact on organisations- Human Resources – Qualifications of employees		The level of qualification of the workforce (degrees) The disciplines and specialities available The experience and expertise of the employees

Impact	Indicator	
	Indicator nominal	Indicator ordinal/numerical
Impact on organisations- Human Resources – Work conditions	Implementing new norms or new equipment related to health and safety Work perspectives	The rate of employee satisfaction towards general work conditions offered by the organisation The amounts invested in training Salaries
Impact on health – Public health – health care		The length of hospitalisation The availability of different types of treatment and medication The satisfaction rate of beneficiaries
Impact on health – Life-expectancy and fertility		The life expectancy at birth and after 65 years of age Fertility rate
Impact on health – Public health	A new prevention program (awareness and immunisation)	The number of individuals benefitting from new prevention programs (awareness and immunisation) The rate of occurrence of contagious/chronic diseases The prevalence of smoking, alcoholism and drug addiction The prevalence of cancer and cardio-vascular diseases
Impact on health – Health system – general costs		Health expenses (in relation to GDP, to government spending expenditure or per inhabitant)
Impact on health – Health system – workforce		The training and expertise of the workforce
Impact on health – Health system – Infrastructure and medical equipment	Medical equipment	The value of investments in infrastructure and new medical equipment The average age of medical equipment
Impact on health – Health system – Products	Approval of medication	The number of new medical protocols
Impact on the environment– Management – natural resources	A plan for the conservation, protection and restoration of species and the ecosystem A bio-diversity plan A plan for the development of resources in a context of sustainable development	

Impact	Indicator	
	Indicator nominal	Indicator ordinal/numerical
Impact on the environment– Management – Environment (pollution)	A surveillance tool for pollution and its causes A method for detection, reductions or elimination of threats related to pollutants The development of anti-pollution norms	
Impact on the environment– Climate and meteorology	A climatic and Metrologic surveillance method A climatic and meteorological model	
Impact on symbolism – Legitimacy/credibility/visibility	Invitation to lead or participate in diverse forums	Level of knowledge of X by Y Level of appreciation de X by Y
Impact on symbolism – Notoriety recognition	A prize, title, promotion/nomination	Market share
Impact on training – curricula	Training programs	
Impact on training – pedagogical tools	Teaching manuals	
Impact on training – Qualifications	Acquired competence	
Impact on training – Insertion into the workforce		The duration of the period between the end studies and the start of the job
Impact on training – Career	Career path	Salary
Impact on training – Use of acquired knowledge	The use of knowledge at work or in daily life	

Godin/Doré (2006: 25-44)

Anhang B: InterviewpartnerInnen

BMF

Mag. Bernhard Schatz

BMVIT

Dr. Rupert Pichler,

Evaluierungsanbieter/ Evaluierungsexperten

Technopolis: Anton Geyer

KMU-Forschung Austria: Sonja Sheikh

WWTF: Klaus Zinöcker

FFG

Michael Binder, Strategieabteilung

Leonhard Jörg, Strategieabteilung

Sabine Mayer, Strategieabteilung

Klaus Schnitzer, Basisprogramme und Strukturprogramme

Emanuel Glenck, Thematische Programme

Johannes Scheer, Thematische Programme

Anhang C: Interviewleitfaden¹⁸

Einstiegsfrage

Wie zufrieden sind sie generell mit den Evaluierungen der FFG-Programme?

Wie beurteilen Sie die österreichische Evaluierungspraxis vor dem Hintergrund internationaler Erfahrungen? Welche Länder könnten ihrer Meinung nach als Good-Practice für Österreich gelten und warum?

Angebot und Nachfrage nach Evaluierung

Welchen Bedürfnissen an Evaluierungen bzw. Aspekten von Evaluation seitens der Förderinstitutionen bzw. Politik wird bislang nur ungenügend durch das Angebot an Evaluierungsleistungen entsprochen?

Die Ausschreibung von Evaluierungen von FFG Programmen sowie die Formulierung der evaluatorischen Fragestellungen (terms of reference) ist Aufgabe des BMVIT. In anderen Interviews äußerten FFG-Mitarbeiter partielle Mängel an den ToR. Insbesondere wurde die mangelnde Einbeziehung der FFG bei der Erstellung der ToR erwähnt. Wie sehen Sie dieses Problem?

Evaluierung der FFG Programme

Die FFG ist aktuell in einer Umstellungsphase von einer programmspezifischen hin zu einer themenspezifischen Förderstrategie. Können Sie uns erörtern welche möglichen Einflüsse hiervon auf zukünftige Evaluierungen zu erwarten sind? (e.g. Portfolio- bzw. Themenevaluierung statt Instrumentenevaluierung)

Die Haushaltsrechtsreform hat u.a. die Einführung einer wirkungsorientierten Haushaltsführung zum Inhalt. Welche Herausforderungen sehen Sie für die gängige Evaluierungspraxis im FTI-Bereich? (Zielkonflikt Interventionslogik vs Outputlogik) Welche bisherigen Evaluierungen von FFG bzw. BMVIT würden Sie als best-practice Evaluierungen bezeichnen? Welche Merkmale dieser Evaluierungen veranlassen Sie zu diesem Urteil? (Erfolgsfaktoren von Evaluierungen)

Gibt es bestimmte Methoden zur Evaluierung von FFG-Programmen, die Sie für besonders nützlich erachten aber systematisch zu wenig angewandt werden? (e.g. quantitative kausale Wirkungsanalyse)

Gibt es ihrer Meinung nach einen Mismatch zwischen eingesetzten Evaluierungsmethoden und Instrumentencharakteristika bei FFG-Programmen?

Sozioökonomischen Wirkungsdimensionen

¹⁸ Nachdem die Interviewpartner aus unterschiedlichen institutionellen Kontexten stammen, wurde der Fragebogen jeweils entsprechend angepasst. Die hier dargestellte Version beinhaltet im Sinne der Mengenlehre die Vereinigungsmenge der Fragen der verschiedenen Fragebögen.

Mit welchen Wirkungsdimensionen gibt es bei bisherigen Programmevaluationen die größten Probleme (das bedeutet: fehlende Daten, Indikatorendefizite,...)?

Wird die Wirkungsdimension soziale Innovation (z.B. Inklusion, Integration, Gender, Diversity, Teilhabefaktoren etc.) in den thematischen Programmen des BMVIT mitgedacht/angestrebt und wurden diese Bereiche bereits evaluiert?

Um die Nachhaltigkeit (im Sinne von nachhaltiger Entwicklung – 3 Säulen: ökonomische Nachhaltigkeit, soziale Nachhaltigkeit und ökologische Nachhaltigkeit) eines Programms zu evaluieren – wo sehen Sie die derzeit bestehenden Lücken? Wo besteht der größte Bedarf? Balance: Welche Säulen sind derzeit am besten messbar; besteht eine Ausgewogenheit bei der Evaluierung der drei Säulen?

Wissen Sie von Wirkungsdimensionen, die im internationalen Bereich Teil in entsprechenden F&E Programmstrategien sind, in Österreich aber noch nicht etabliert sind?

Spontan, was würden Sie sagen, wenn zusätzlich zu sozioökonomischen Wirkungsdimensionen auch soziokulturelle Wirkungsdimensionen (wie z.B. Veränderungen in Bezug auf gesellschaftliche Werte, Normen; Ethik, Moraletik, Toleranz,...) angedacht würden?

Erwartungen an die Studie

Unsere Studie kann als eine Art Meta-Studie zum Thema Evaluierung verstanden werden. Was würden Sie aus der Perspektive der Förderpraxis als wichtige Inhalte bzw. Fragestellungen die im Rahmen dieser Studie bearbeitet werden sollten, erachten?

Unsere Studie soll als ein wesentliches Ergebnis Hinweise auf eine optimierte Anpassung von Evaluierungsansätzen an die Programme der FFG und des BMVIT beinhalten. Wo sehen Sie denn aktuell Probleme bzw. das größte Verbesserungspotenzial? Wirkungsanalysen präsentiert.



GFF – Gesellschaft zur Förderung der Forschung

c/o AIT Austrian Institute of Technology

2444 Seibersdorf

Tel. +43 50 550 2040

Fax +43 50 550 2044

eMail barbara.dagott@nes.at

