

4 Monetarismus II: rationales Erwartungsgleichgewicht

4-1 Zwei Grundannahmen des Monetarismus II

4-1-1 Die rationalen Erwartungen

Lucas (1996) identifiziert die eigentliche Forschungsaufgabe monetärer Theorie mit der Lösung des Hume-Problems: Die Geldmengenveränderungen seien zwar neutrale Einheitsveränderungen, sie würden jedoch Bewegungen von Beschäftigung und Produktion in die gleiche Richtung induzieren (Lucas 1996: 644). Lucas und die Neuklassik haben dieses Problem dahingehend thematisiert, im Rahmen des allgemeinen Gleichgewichtsmodells den Übergang von einer quantitativtheoretisch bestimmten Gleichgewichtsposition zu einer anderen zu erklären. Damit war eine Verlagerung des analytischen Schwerpunkts vom komparativ-statischen auf das dynamische Modell einhergegangen, in deren Zuge das intertemporale Entscheidungsproblem und die Erwartungen zunehmend an der Bedeutung gewonnen haben.

Wie im vorhergehenden Kapitel dargestellt, hat Friedman diesbezüglich – zusammen mit asymmetrischer Informationsverarbeitung – die adaptiven Erwartungen postuliert, durch die Geldillusion erlaubt wird. Bei adaptiven Erwartungen richten sich die Erwartungen ausschließlich an den Vergangenheitswerten bestimmter Variablen aus. Diese „backward-looking“-Hypothese weist jedoch eine Schwäche auf, weil sie einen systematischen Schätzfehler der Wirtschaftssubjekte unterstellt, in dem Sinne, dass die Wirtschaftssubjekte wiederholt falsche Erwartungen bilden. Diese Schwäche bezieht sich vor allem auf „(i) the assumption that economic agents only partially adjust their expectations by a fraction of the last error made; and (ii) the failure of agents to take into consideration additional information available to them other than past values of the variable concerned, despite making repeated errors“ (Snowdon u.a. 1994: 191).

Die neuklassische Ökonomie hat an diesem Punkt angesetzt und stattdessen die rationale Erwartungshypothese postuliert, die auf Muth (1961) zurückgeht. Im Gegensatz zur Forschungspraxis seiner zeitgenössischen Ökonomie, in der verschiedene Erwartungsformeln im Kontext des dynamischen Modells als „free parameter“ von außen eingeführt wurden, hat Muth seine alternative Erwartungshypothese aufgestellt:

„I should like to suggest, that expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory. At the risk of confusing this purely descriptive hypothesis with a pronouncement as to what firms ought to do, we call such expectations ‘rational.’”
(Muth 1961: 316).

Die Muth’sche rationale Erwartungshypothese, die von der neuklassischen Ökonomie in das makroökonomische Modell inkorporiert wurde, behauptet, dass die Erwartungen entsprechend der Gleichgewichtslösung des Modells gebildet werden. Erwartungen werden somit vom Modell endogen erklärt. Die Grundzüge der rationalen Erwartungen lassen sich anhand des Beispiels der Inflationserwartungen der Wirtschaftssubjekte verdeutlichen:

$$\pi_{t+j}^e = E(\pi_{t+j} | \Omega_t) \quad (4.1)$$

$$\pi_{t+1} - \pi_{t+1}^e = \varepsilon_{t+1} \quad (4.2)$$

wobei π_{t+j} die tatsächliche Inflationsrate in der Periode $t+j$, π_{t+j}^e die erwartete Inflationsrate für die Periode $t+j$, ε_t den Prognosefehler bezeichnet, der einen Mittelwert von Null hat und mit dem zum Prognosezeitpunkt verfügbaren Informationsset Ω_t nicht korreliert ist.

- In die Erwartungsbildung gehen die künftigen und damit zu prognostizierenden Ereignisse, nicht die vergangenen Entwicklungen, ein. Die rationalen Erwartungen sind vorwärtsblickend („forward-looking“).
- Die Annahme (4.1) ist die Spezifikation der rationalen Erwartungshypothese, dass die subjektiven Erwartungen („Prognosen“) der Wirtschaftssubjekte mit dem objektiven Erwartungswert identisch sind. $E(\cdot)$ stellt die mathematische (objektive) Erwartung über die Inflationsrate für die Periode $t+j$ dar, die unter Zuhilfenahme der zum Prognosezeitpunkt t verfügbaren Informationsmenge Ω_t kalkuliert wird. Die so gebildeten Erwartungen sind rational, weil die Wirtschaftssubjekte alle verfügbaren, für die Prognose einer bestimmten Variablen relevanten Informationen benutzen. Die Annahme, dass die objektive Wahrscheinlichkeitsverteilung und ihr Mittelwert durch das ökonomische Modell selbst generiert werden, läuft darauf hinaus, dass die (subjektiven) Erwartungen mit der optimalen Prognose des zugrunde liegenden ökonomischen Modells („true model“) zusammenfallen.
- Die Annahme (4.2) weist darauf hin, dass die rationalen Erwartungen keine vollständige Prognose unterstellen. Sie erlauben einen Prognosefehler, der jedoch auf exogenen, unvor-

hershkbaren, zufällig auftretenden Schocks beruht. Der Prognosefehler ist jedoch im Durchschnitt Null: $E[\pi_{t+j} - E(\pi_{t+j} | \Omega_t)] = 0$. Das bedeutet, dass der Prognosefehler keine systematische Komponente beinhaltet. Die rationalen Erwartungen produzieren unverzerrte Schätzung der stochastischen ökonomischen Prozesse, die zu antizipieren sind.

Kurzum unterstellt die rationale Erwartungshypothese, dass die Wirtschaftssubjekte (i) das korrekte Modell (in diesem Fall die neoklassische Theorie der Interdependenz zwischen Geldmenge und Preisniveau) kennen und (ii) sich auch in Übereinstimmung mit ihm verhalten. Die Wirtschaftssubjekte können nur aufgrund der unantizipierbaren Schockvariable „getäuscht“ werden, aber solche Erwartungsfehler gleichen sich langfristig aus.

Die rationale Erwartungshypothese weist einige positive Aspekte auf: Beseitigung der ad hoc Annahme bezüglich der Erwartungsbildung, Beseitigung des systematischen Erwartungsfehlers, Nutzung aller möglichen Informationen nach dem Muster des mikroökonomischen Optimierungskalküls, Einbezug des makroökonomischen Wissens in die ökonomischen Handlungen, dessen weitere Implikation darin besteht, dass das Verhalten der Wirtschaftssubjekte von den wirtschaftspolitischen Maßnahmen nicht unberührt bleibt. Diese Ansicht bildet die Begründungsmöglichkeit der sog. Lucas-Kritik, zu der wir später kommen werden. Ohne Zweifel bleibt der neoklassischen Ökonomie das Verdienst, dass sie die Bedeutung der Erwartungsbildung für die makroökonomische Theorie und für die Wirtschaftspolitik in analytisch erfassbarer Form herausgearbeitet hat.

Trotz dieser positiven Aspekte ist die rationale Erwartungshypothese neoklassischer Prägung einer Reihe von ernsthaften Einwänden ausgesetzt. Hinzuweisen ist auf die Kosten der Informationsbeschaffung. Sie können zum Verzicht auf die rationalen Erwartungen zugunsten der adaptiven Erwartungen führen. Gravierender ist jedoch die Annahme, dass die Wirtschaftssubjekte das wahre Modell der Ökonomie kennen und sich demnach verhalten. Aber wie Riese (1983b: 256) aufzeigt, stützt sich diese Annahme auf zwei Prämissen: eine kollektive Verarbeitung der Informationen und eine kollektive Informationsverwendung nach dem wahren Modell, worunter die neoklassische Ökonomie die Quantitätstheorie versteht. Nach Riese (1983b: 258f.) bleibt die erste Prämisse unbefriedigend, weil zwischen „neoklassisch“ und „keynesianisch“ handelnden Wirtschaftssubjekten zu unterscheiden ist; die zweite Prämisse ist nicht stichhaltig, weil sie gleichförmige Interessen der Wirtschaftssubjekte voraussetzt. „Even if, as the model-builder may assume, the agents in the economy believe the model itself, this consistency does not guarantee that the agents agree with each other regarding the nature and probabilities among various future shocks (...) No process of expectation formation can achieve equivalent coordination“ (Tobin 1980: 25f.). Mithin ist ein Pro-

zess der Konvergenz der Erwartungen auf Basis subjektiver zu objektiven Wahrscheinlichkeiten ausgeschlossen.

4-1-2 Beständige Markträumung

Die zweite Grundannahme der neoklassischen Ökonomie ist, dass die Ökonomie anhand der beständigen Markträumung beschrieben werden kann: Die Märkte werden immer in Übereinstimmung mit dem walrasianischen Gleichgewichtmodell geräumt und befinden sich im kontinuierlichen – sowohl kurz- als auch langfristigen – Gleichgewicht.⁴² Das Gleichgewicht wird hier ausschließlich als Markträumung begriffen: Keine Überschussnachfrage, keine Rationierung.

Beständige Markträumung heißt, dass sich Preise und Mengen in jeder Situation aus den Optimierungskalkülen der Wirtschaftssubjekte ergeben. Es kann wohl bei exogenen Schocks die Abweichungen vom natürlichen Niveau geben. Die neoklassische Ökonomie deutet sie jedoch nicht als Ungleichgewicht, sondern als eine Kontinuität des Gleichgewichts. Demnach führt jede Veränderung der Produktivität, der Präferenzen oder auch der Geldmenge nicht zum Cobweb-Prozess, sondern sofort zu einer neuen Gleichgewichtskonstellation, d.h. die Märkte springen als Reaktion auf Veränderungen exogener Variablen von einem Gleichgewichtspunkt zum anderen.⁴³

Die Annahme der beständigen Markträumung läuft darauf hinaus, dass es in der neoklassischen Ökonomie auch temporär keine unfreiwillige Arbeitslosigkeit gibt: Wir haben immer Vollbeschäftigung – gleichgültig wie hoch das Beschäftigungsniveau ist und wie hoch die Arbeitslosenquote ist. Beschäftigungsschwankungen gehen auf Änderungen in Präferenzen bzw. in der Produktivität zurück. In dieser Vorstellung ist die Arbeitslosigkeit ein mikroökonomisches Phänomen in dem Sinne, dass sich die Individuen gegen ein Arbeitsangebot und zugunsten der Freizeit entscheiden.

⁴² Tobin (1980: 33) bezeichnet dies als magna carta, die Leon Walras der Ökonomik gegeben hat.

⁴³ „Sofortige Markträumung und rationale Erwartungen sind für die Neoklassik zwei Seite der gleichen Medaille. Denn wenn alle Wirtschaftssubjekte über vollständige Informationen verfügen und das ‚wahre‘ Modell zur Erklärung der Ökonomie kennen, dann wird jede neue Information zur sofortigen Marktreaktion führen. Alle Wirtschaftssubjekte passen sich bei neuen Informationen aufgrund nun veränderter Bedingungen an und springen in ein neues Gleichgewicht“ (Heine/Herr 2003: 297).

Mit Betz (1993: 37) lässt sich feststellen, dass die Gleichgewichtskriterien paradigmaabhängig sind. Die Marktträumung als Gleichgewichtskriterium ergibt sich deshalb, „weil die Akteure a) die Preise parametrisch behandeln und diese Preise die einzigen Informationen sind, die in die Planung ihrer Handlungen eingehen, weil sie b) erwarten (...), dass sie zu den gegebenen Preisen beliebige Mengen kaufen bzw. verkaufen können (...) sowie c) weil ihre (...) Handlungsmöglichkeiten, auf die Äußerung eines notional demand, die Bekanntgabe von Mengenplänen, festgelegt sind“ (Betz 1993: 26f.).

4-2 Unvollkommene Information und Lucas-Angebotsfunktion

Wir behandeln in diesem Abschnitt, wie Lucas das eingangs erwähnte Hume-Problem erklärt. Seine Erklärung stützt sich auf das quantitätstheoretische Neutralitätspostulat und das Allgemeine Gleichgewichtsmodell. Sein Vorgehen hebt sich jedoch vom traditionellen Monetarismus in einem Punkt ab: Während der traditionelle Monetarismus, so Lucas (1996: 668ff.), den dynamischen Anpassungsprozess als außerhalb des allgemeinen Gleichgewichts liegend betrachtet und daher die Dynamik einen patched-in Charakter behält, fasst Lucas ihn als Folge des intertemporalen Entscheidungsproblems und damit als Kontinuum von Gleichgewichten auf. Auf diesem methodischen Standpunkt präsentiert er seine Angebotsfunktion, die den Zusammenhang von Inflationsrate und Outputniveau erklärt.⁴⁴

Lucas schließt in seiner Ableitung der Angebotsfunktion alle Formen der Geldillusion aus und unterstellt, dass die Preise marktträumend und die Erwartungen rational sind. Der Kernpunkt dabei ist die Annahme unvollkommener Information. Die Idee ist wie folgt (vgl. Lucas 1972: 103f.): Die entscheidende Informationsquelle, die die Allokationsentscheidung der Wirtschaftssubjekte bestimmt, sind die Preise auf dem Markt, auf dem das jeweilige Wirtschaftssubjekt agiert. Das aggregierte Angebot hängt dabei von den relativen Preisen ab. Das Problem ist jedoch, dass die Akteure sich in einer Situation befinden, in der man relative Preisänderungen nicht von Änderungen des Preisniveaus unterscheiden kann: Sie beobachten zwar die Preisänderungen in ihrem Markt, wissen

⁴⁴ Für die weitere Diskussion ist angebracht zu bemerken: Es handelt sich um Okuns Gesetz (Okun 1962), das einen negativen Zusammenhang zwischen Arbeitslosenquote und realem BIP formuliert. Nach Okun ist eine Abnahme der Arbeitslosigkeit um 1% mit einem zusätzlichen Wachstum des realen BIP um ca. 2,5% verbunden. Dem Begriff nach entspricht das normale bzw. potentielle Outputniveau der natürlichen Arbeitslosenquote.

jedoch nicht, ob diese Preisänderungen reale (d.h. Änderung in relativen Preisen) oder monetäre (d.h. Änderung des allgemeinen Preisniveaus) Ursachen haben. Denn die Preise liefern annahmegemäß die relevante Information unvollständig. Angesichts dieses „signal extraction“-Problems werden die Wirtschaftssubjekte gezwungen, auf die Preisänderungen in intermediärer Weise zu reagieren. Diese Reaktion lässt die Preise und den Output sich in gleicher Richtung ändern, allerdings in geringerem Ausmaß als bei einer bekannten relativen Nachfrageverschiebung.

Um die so skizzierte Idee zu formalisieren, benutzt Lucas als analytische Instrumente Samuelsons Overlapping-Generations-Ökonomie in Verbindung mit Phelps Inselparabel. Das erste bietet den Rahmen, die intertemporalen Entscheidungsprobleme zu erfassen, wobei das Geld als Wertaufbewahrungsmittel gefasst wird; das zweite erlaubt ein Modell mit lokalen Märkten, um die unvollständige Information als wichtige Ursache der Friktionen zu modellieren.

Stellen wir Lucas' „Inselökonomie“ formal dar. Sie besteht aus n Konkurrenzmärkten, auf denen die ein identisches Produkt herstellenden, individuellen Anbieter stochastisch verteilt werden.⁴⁵ Mit dieser Annahme ist das Aggregationsproblem ausgeschlossen. Weiterhin wird angenommen, dass die Nachfrage nach den Gütern in jeder Periode ungleichmäßig verteilt ist. Somit lässt sich der Zustand eines Marktes durch zwei Faktoren beschreiben: Erstens durch einen Nachfrageschock z_t , der seriell unabhängig ist und den Mittelwert Null sowie eine endliche Varianz besitzt; zweitens durch den Wert des um eine Periode verzögerten Angebots y_{t-1} , der die spezifische Allokation an jedem Markt darstellt. Die Annahme von rationalen Erwartungen besagt, dass die Wirtschaftssubjekte die ersten und zweiten Momente aller Wahrscheinlichkeitsverteilungen kennen mögen. Mithilfe dieser Information können sie eine lineare Erwartung für die Zufallsvariablen bilden, die sie nicht kennen. Die Angebotsmenge für den repräsentativen Markt z im Zeitpunkt t besteht aus einer „normalen“ Komponente y_t^* , die der natürlichen Arbeitslosenquote entspricht, und aus einer zyklischen Komponente $y_t^c(z)$, die eine Abweichung vom normalen Outputniveau darstellt.⁴⁶

$$y_t(z) = y_t^* + y_t^c(z) \quad (4.3)$$

⁴⁵ Die Produzenten und die Arbeiter sind hier nicht zu unterscheiden, da es sich um das „yeoman farmer“-Modell handelt, in dem die Haushalte als Anbieter an Gütern sowie an Arbeit angenommen werden. Als Referenz nehmen wir die Darstellung in Lucas (1973). Vgl. Auch Lucas (1972), Sargent (1994: 332ff.), Leslie (1993: 136ff.).

⁴⁶ Folgende Gleichungen werden in logarithmierter Form dargestellt.

Da die normale Komponente vorgegeben ist, verbleibt, die zyklischen Komponente des Angebots zu bestimmen, die folgende Form annimmt:

$$y_t^c(z) = \gamma \left[p_t(z) - E(p_t | \Omega_t(z)) \right] + \lambda y_{t-1}(z) \quad (4.4)$$

mit $p_t(z)$ als gegenwärtigem Preis zum Zeitpunkt t auf dem z -Markt und $E[p_t | \Omega_t(z)]$ als Erwartung für das Preisniveau aufgrund der zum Zeitpunkt t auf dem z -Markt verfügbaren Information $\Omega_t(z)$ sowie $|\lambda| < 1$.

Die Gleichung (4.4) besagt, dass die Wirtschaftssubjekte positiv auf die von ihnen wahrgenommene Änderung des relativen Preises reagieren. Eine Erklärung dieser Gleichung wird durch die Idee der intertemporalen Substitution gegeben – nämlich nach dem Motto „making hay while the sun shines“: Mehr Arbeiten und mehr Produzieren, wenn die relativen Preise günstig sind; mehr Freizeit und weniger Produzieren, wenn die relativen Preise ungünstig sind (Leslie 1993: 137).⁴⁷

Angenommen wird, dass es zwei Quellen der Preisstörungen gibt: Die erste kommt aus dem monetären Faktor, der lediglich auf das allgemeine Preisniveau wirkt; die zweite ist die für den z -Markt spezifische Wirkung des relativen Preises. Der repräsentative Anbieter (an Gütern bzw. an Arbeit) ist nun mit dem Problem der Signalentnahme konfrontiert. Er hat anhand der zum Zeitpunkt t auf seinem Markt verfügbaren Information zu entscheiden, ob die gegenwärtige Preisänderung durch eine Änderung im Preisniveau oder im relativen Preis verursacht wurde. Folglich ist es für die Lösung notwendig, die Informationsmenge $\Omega_t(z)$ zu spezifizieren. Nach Lucas sind zwei Informationsquellen relevant. Erstens beinhaltet $\Omega_t(z)$ eine Information, welche eine Reihe von verzögerten Variablen auf allen Märkten umfasst. Daraus kann a priori die Verteilung des Preisniveaus, also der Mittelwert \bar{p} und die Varianz σ^2 , deduziert werden. Zweitens gilt der gegenwärtige Preis $p_t(z)$ auf dem z -Markt, der vom allgemeinen Preisniveau um z abweicht (diese Kennzeichnung nehmen wir von Lucas. Er kennzeichnet der Vereinfachung halber den Markt durch diese Differenz): z besitzt den Mittelwert Null (weil $\sum_1^n z = 0$ ist) und die Varianz τ^2 . Der Preis am z -Markt zum Zeitpunkt t ist dann

⁴⁷ Lucas (1972) hat eine Ableitung dieser Gleichungen aus dem individuellen Nutzenmaximierungsproblem geliefert, deren Verständnis ein hohes mathematisches Vermögen beansprucht. Eine Darstellung in einfacher mathematischer Form findet sich in Romer (2001: 266ff).

$$p_t(z) = p_t + z \quad (4.5)$$

In Bezug auf den Preis $p_t(z)$ ist erstens zu bemerken, dass er als eine Funktion des möglichen Zustandes des Marktes z aufgefasst wird: „This characterization permits a treatment of the relation of information to expectation (...)“ (Lucas 1972: 104). Zweitens demonstriert Lucas (Lucas 1972: 111-113), dass er der markträumende, einzige gleichgewichtige Preis ist, zusammen mit dem korrespondierenden Output also die einzige Lösung des Modell ist.⁴⁸ Dieses Ergebnis beruht allerdings auf der Annahme, dass der Anbieter an Gütern und Arbeit auf seiner Angebotsfunktion ist: In Clowers termini (Clower 1963) handelt es sich um das „notional“ Angebot, das auf die „notional“ Nachfrage trifft. Mithin existiert im Lucas’ Modell keine (unfreiwillige oder besser: marktbedingte) Unterbeschäftigung.

Es kommt nun auf die Prognose des allgemeinen Preisniveaus p_t an, das nicht direkt zu beobachten ist, und zwar mithilfe der so spezifizierten Information $\Omega_t(z)$. Es gilt:⁴⁹

$$E[p_t | I_t(z)] = E[p_t | p_t(z), \bar{p}_t] = (1 - \theta) p_t(z) + \theta \bar{p}_t \quad (4.6)$$

wobei der Parameter $\theta = \tau^2 / (\sigma^2 + \tau^2)$ den Anteil der Varianz von $p_t(z)$ darstellt.⁵⁰ Die rationale Prognose von p_t ergibt sich als ein gewichteter Durchschnitt von gegenwärtigem Preis und Mittelwert des Preisniveaus. Die Kombination der Gleichungen (4.3), (4.4) und (4.6) ermittelt die Angebotsfunktion für den repräsentativen Markt z

⁴⁸ Von Lucas zu zitieren: „One would like a theory which accounts for the observed movements in quantities (...) as an optimizing response to observed movements in prices“

⁴⁹ Für die schrittweise Ableitung vgl. Sargent (1994: 334f.). Eine wichtige technische Methode ist dabei die rekursive Methode. Über ihre Bedeutung schreibt Sargent (1996: 543): „Virtually every macroeconomist today reads, writes, and thinks in terms of dynamic programming. Recursive reasoning unearthed the ‘time-consistency’ problem, and led to characterizing the types of government policies that could be implemented sequentially.“

⁵⁰ Man kann die Preisänderung aufgrund der un konditionalen Erwartungen des Preisniveaus, nämlich \bar{p}_t , schätzen. Da jedoch gefordert wird, dass die Wirtschaftssubjekte alle Information effizient benutzen sollen, geht es hier um die konditionale Prognose. Diese Denkrichtung ist nachzuvollziehen, wenn man bedenkt: Nehmen wir an, dass die Schwankungen des Preisniveaus, also die Varianz σ^2 , dreifach höher als die Schwankungen des relativen Preises τ^2 sind. Dann kann man sich vorstellen, dass die beobachtende Preisänderung meistens – in diesem Fall in drei von vier Fällen – durch Änderungen im Preisniveau ausgelöst wird. Insofern liefert die relative Varianz eine Gewichtungsmethode, um den konditionalen Mittelwert abzuleiten.

$$y_t(z) = y_t^* + \theta\gamma(p_t(z) - \bar{p}_t) + \lambda y_{t-1} \quad (4.7)$$

Aus dem Durchschnitt der einzelnen Angebotsfunktion geht die aggregierte Angebotsfunktion hervor

$$y_t = y_t^* + \theta\gamma(p_t - \bar{p}_t) + \lambda(y_{t-1} - y_{t-1}^*) \quad (4.8)$$

Die Gleichung (4.8) ist die Lucas-Angebotsfunktion, also eine neuklassische Variante der Phillipskurve, wobei der dritte Term der Gleichung die Persistenz der vergangenen Allokation darstellt.⁵¹ Freilich kann diese Gleichung für die Darstellung des Zusammenhangs von Inflationsrate und Outputniveau umformuliert werden, indem man das um eine Periode verzögerte Preisniveau p_{t-1} addiert und wieder abzieht:

$$\begin{aligned} y_t &= y_t^* + \theta\gamma((p_t - p_{t-1}) - (\bar{p}_t - \bar{p}_{t-1})) + \lambda(y_{t-1} - y_{t-1}^*) \\ &= y_t^* + \theta\gamma(\pi_t - \pi_t^e) + \lambda(y_{t-1} - y_{t-1}^*) \end{aligned} \quad (4.9)$$

bzw.

$$\pi_t = \pi_t^e + (\theta\gamma)^{-1} y_t^* - (\theta\gamma)^{-1} \lambda \hat{y}_{t-1} \quad (4.10)$$

wobei \hat{y}_t die lineare Abweichung des Angebots vom natürlichen Niveau im Zeitpunkt t darstellt. Die Inflationsrate hängt ab von der erwarteten Inflationsrate und dem Outputniveau. Man erhält damit eine Phillipskurve.

Die Hauptresultate, die sich aus der so skizzierten Angebotsfunktion ergeben, sind folgende: Erstens betrachtet die Lucas-Angebotsfunktion die natürliche Arbeitslosenquote als eine Systemeigenschaft. Da die rationalen Erwartungen keinen systematischen Prognosefehler implizieren, also $E[p_t - E(p_t | p_t(z), \bar{p}_t)] = E[p_t - \bar{p}_t] = 0$, wird der aggregierte Output ohne monetäre Schocks durch das natürliche Outputniveau (oder seine Geschichte) bestimmt: Der aggregierte Output folgt dem Pfad der natürlichen Arbeitslosenquote, welcher unabhängig von den nominalen Preisen ist. Das monetaristische Neutralitätspostulat wird damit bestätigt.

⁵¹ Leslie (1993: 142) weist darauf hin, dass dieser Term zwar eine minimale Begründung für die in der Realität feststellbare Persistenz, aber eine ad hoc Konstruktion ist.

Zweitens sind reale Wirkungen von der antizipierten und nichtantizipierten Geldmengenveränderung zu unterscheiden (vgl. Barro 1975, 1976). Aggregierte Nachfrageschocks, die hauptsächlich aus nichtantizipierten Änderungen im Geldangebot resultieren, bewirken den Erwartungsfehler, ausgedrückt in $(p_t - \bar{p}_t)$ bzw. $(\pi_t - \pi_t^e)$, und damit die Abweichungen des Outputs von seinem natürlichen Niveau, das sich unter vollständiger Information ergeben würde. „These deviations occur because agents are obliged to infer current general price movements on the basis of incomplete information“ (Lucas 1976: 122). Sofern die Geldpolitik in einer systematischen Form implementiert wird, hat sie jedoch keine Wirkungen auf die natürliche Arbeitslosenquote, und zwar auch in der kurzen Frist. Denn die rational agierenden Wirtschaftssubjekte können mithilfe des „wahren“ Modells die Konsequenz der Geldpolitik simulieren.⁵²

Es gilt hier, auf einen wichtigen Unterschied zwischen Friedmans und Lucas' Modell hinzuweisen. Es handelt sich um die Eigenschaft des Zusammenhangs zwischen dem tatsächlichen und erwarteten Preisniveau (bzw. der Inflationsrate). Friedmans Modell unterstellt aufgrund der adaptiven Erwartungen eine Zeitverzögerung zwischen Erwartung und Realisierung der Inflation. Im Kontrast dazu postuliert Lucas' Modell, dass die „Rigidität“, die das kurzfristige Angebotsverhalten dominiert, aus der unvollkommenen Information resultiert. Lucas' Modell „formulated a version of Friedman and Phelps's natural rate theory that was consistent with the new equilibrium concept, and displaced the older distinction between short and long run in favor of one between expected and unexpected outcomes“ (Sargent 1996: 539; vgl. auch McCallum 1989: 187).

Die Logik, dass die monetären Veränderungen realen Effekt haben, nur weil die Wirtschaftssubjekte zwischen der realen und monetären Bewegung nur unvollkommen unterscheiden können, heißt drittens, dass die Existenz des Trade-off eine Folge des rationalen Verhaltens der Wirtschaftssubjekte ist, die mit der Unsicherheit über die Preisänderungen konfrontiert sind. Die Phillipskurve wird nicht mehr als ein wirtschaftspolitisch ausbeutbarer Zusammenhang gesehen, sondern als eine zentrale Eigenschaft der Lösung zu einem allgemeinen Gleichgewichtssystem erklärt (Lucas 1972: 122).

Viertens ist das Verhältnis zwischen dem Neigungsparameter der Phillipskurve $\theta\gamma$ und der Varianz von Preiskomponenten zu erklären. Ein kleines τ^2 bedeutet, dass die nichtantizipierte Preis-

⁵² Es ist nochmals zu betonen, dass die neoklassische Ökonomie die Geldpolitik als Geldmengenpolitik versteht und auf dem quantitativ-theoretischen Neutralitätspostulat besteht. Das „wahre“ Modell wird damit allgemein als ihre eigene Theorie und speziell als Quantitätstheorie interpretiert.

bewegung faktisch die Änderung des allgemeinen Preisniveaus reflektiert. In diesem Fall reagiert das Angebot nicht auf die Preisänderung. Die Angebotskurve wird mithin vertikal. Im Fall eines kleinen σ^2 , was ein stabile Preisniveau bedeutet, löst dagegen die Preisänderung eine große Reaktion des aggregierten Angebots aus, weil sie als Änderung zugunsten des relativen Preises gedeutet wird.

Dieser Umstand zeigt weiterhin auf, dass der Neigungsparameter der Phillipskurve $\theta\gamma$ in einem inversen Verhältnis mit der monetären Varianz σ^2 steht, so dass $\theta \rightarrow 0$ wird, wenn $\sigma^2 \rightarrow \infty$ (Barro 1976).⁵³ Das impliziert, dass, wenn die monetären Instanzen immer wieder versuchen, den Trade-off auszubeuten, sich die Varianz σ^2 im Vergleich zu τ^2 vergrößert und entsprechend die privaten Akteure umso unsensibler auf Nachfrageänderungen reagieren werden. Denn dadurch wird die Signalfunktion des Preises ineffektiv. Aus diesem Grunde plädieren Lucas und die neuklassische Ökonomie für die starren Geldmengenregeln und lehnen den Versuch der Nachfragesteuerung strikt ab.⁵⁴

Eine weitere Implikation ergibt sich daraus, dass die Änderung der relativen Gewichte von σ^2 und τ^2 als Reaktion auf die monetäre Störung den Wert des Parameters $\theta\gamma$ verändert. Das heißt, die Neigung der Phillipskurve bleibt nicht konstant für die Politikveränderungen. Dieser Umstand demonstriert, „wie sich die optimalen Entscheidungsregeln der Wirtschaftssubjekte wandeln, wenn es bei den Zufallsprozessen jener exogenen Variablen, auf die sich ihre Entscheidungen stützen, zu Veränderungen kommt“ (Sargent 1994: 337). Die sog. Lucas-Kritik, die unten zu diskutieren sein wird, setzt an diesem Punkt an.

Nicht nur untermauert die Lucas-Angebotsfunktion die monetaristischen Argumente, sondern sie radikalisiert sie zugleich. Bezogen auf die Phillipskurven-Diskussion wurde behauptet, dass bei rationalen Erwartungen weder eine kurzfristige noch eine langfristige Phillipskurve existiert: Auch kurzfristig ist antizipierte Geldpolitik neutral. Die realen Wirkungen der Geldpolitik beruhen ledig-

⁵³ Lucas (1996: 678): „The signal processing feature of the model of Lucas (1972) implied that the magnitude of a money multiplier should decline as the variance of money changes increase.“

⁵⁴ „(T)he choice of a monetary policy is equivalent to the choice of a density function f governing the stochastic rate of monetary expansion. Densities f which are concentrated on a single point correspond to fixing the rate of monetary growth at a constant percentage rate k . Following Friedman, we shall call such a policy a *k-percent rule*“
Gleichzeitig behauptet Lucas, „(...) that if a k-percent rule is followed the competitive allocation will be Pareto-optimal“ (Lucas 1972: 119).

lich auf Prognosefehlern. Mithin hat die neuklassische Ökonomie das Verhältnis zwischen Lang- und Kurzfristigkeit im Monetarismus I durch das Verhältnis zwischen Antizipierbarkeit und Prognosefehler ersetzt. Die Existenz des Trade-off wurde jedoch als Gleichgewichtslösung in der Umwelt, wo die Informationen unvollständig sind, hergeleitet und daher als wirtschaftspolitisch nicht ausbeutbar betrachtet. Welche wirtschaftspolitischen Implikationen die Lucas-Angebotsfunktion mit rationalen Erwartungen beinhaltet, beleuchten wir im nächsten Abschnitt.

Die starke Implikation der Lucas-Angebotsfunktion, allein die unantizipierten monetären Schocks hätten irgendwelche Effekte auf realwirtschaftliche Aktivitäten, ist nicht stichhaltig. Sie beruht auf der eigenartigen Interpretation der rationalen Erwartungen (die Annahme des neuklassischen Monetarismus als *das* wahre Modell) und dem Begreifen des Gleichgewichts als Markträumung. Woodford (2003: 173ff.) macht darauf aufmerksam, dass die neuklassische Feststellung im Widerspruch zu den Effekten stehe, die in der „structural VAR(Vector Auto-Regression)“-Literatur identifiziert wurden. Diese kommen zu dem Ergebnis, dass es einen substantiellen Effekt gibt und dass darüber hinaus der Effekt „long-lasting“ ist.

4-3 Wirtschaftspolitische Implikationen

4-3-1 Das neuklassische Makromodell

Um die wirtschaftspolitischen Implikationen der neuklassischen Ökonomie zu diskutieren, stellen wir ein stochastisches Makromodell in einfacher Form dar, das aus folgenden drei log-lineare Gleichungen (= Strukturgleichungen) besteht:⁵⁵

$$y_t = y_t^* + \beta(p_t - E_{t-1}(p_t)) + u_t \quad (4.11)$$

$$y_t = m_t - p_t + v_t \quad (4.12)$$

$$m_t = b_0 + b_1 y_{t-1} + \varepsilon_t \quad (4.13)$$

⁵⁵ Diese Modellstruktur bezieht sich auf Sargent/Wallace (1975) und Sheffrin (1996: 35ff).

Gleichung (4.11) ist die oben skizzierte aggregierte Angebotsfunktion, wobei u_t der zufallsverteilte Störterm mit einem Mittelwert von Null und mit endlicher Varianz ist. Er steht für unerwartete Änderungen in Produktivität, Präferenzen etc., die die Angebotskurve verschieben. Und $E_{t-1}(p_t)$ bezeichnet die Erwartungen für das Preisniveau zum Zeitpunkt t , die zum Zeitpunkt $t-1$ gebildet wurden. y_t^* stellt das natürliche Outputniveau dar und wird der Einfachheit halber als konstant angenommen.

Das zweite Element des Modells ist die aggregierte Nachfragefunktion (4.12), wobei v_t der nachfrageverschiebende Zufallsterm ist, der den Mittelwert Null und eine endliche Varianz besitzt. Er kann auch als derjenige betrachtet werden, der die Nachfragestörungen jenseits der Geldmengenänderungen auffängt. Die aggregierte Nachfragefunktion geht hier auf die Quantitätsgleichung zurück.⁵⁶ Wichtig ist in dieser aggregierten Nachfragefunktion das inverse Verhältnis zwischen Preisniveau und Output. Das Geld geht in das Modell durch seinen (Realkassen)Effekt auf die aggregierte Nachfrage ein.

Das dritte Element des Modells, Gleichung (4.13), beschreibt das Verhalten der Geldpolitik. Hier ist angenommen, dass die Geldpolitik auf die Verringerung der Schwankungen im Output zielt. Zwar ist diese Charakterisierung der Geldpolitik nicht das neuklassische Konzept, aber wir nehmen dies für die Vorbereitung der späteren Diskussion an. Nach Gleichung (4.13) bestimmt sich das Geldangebot im Zeitpunkt t , m_t , in Abhängigkeit vom Outputniveau der letzten Periode $b_1 y_{t-1}$ – dieser Bezug lässt von einer Rückkopplungspolitik sprechen – und dem „drift“ b_0 sowie von der Zufallsvariablen ε_t , die formal „white noise“ ist. Der Teil $b_0 + b_1 y_{t-1}$ stellt die systematische Komponente der Geldpolitik dar, in dem Sinne, dass die Parameter b_0 und b_1 durch die monetären Autoritäten gesetzt werden. Dieser Teil ist den Wirtschaftssubjekten bekannt. Die rationalen Erwartungen implizieren, dass die Wirtschaftssubjekte die Rückkopplungsregeln der monetären Autoritäten kennen und sie bei ihrer Erwartungsbildung in Rechnung stellen. Dagegen bildet der ε_t die unsystematische Komponente der Geldpolitik und stellt die geldpolitische Überraschung dar, die die Abweichungen des (tatsächlichen) Outputs vom natürlichen oder potentiellen Niveau bewirkt.

⁵⁶ Alternativ zur Gleichung (4.12) kann man freilich die aggregierte Nachfragefunktion aus der IS-LM-Relation ableiten. In diesem Fall ändert sich auch nichts am Ergebnis.

Das gleichgewichtige Preisniveau und der gleichgewichtige Output ergeben sich aus der Bedingung der Marktträumung, also aus der Gleichsetzung von Angebot und Nachfrage. Unter Berücksichtigung von $E_{t-1}(p_t) = E_{t-1}(m_t) - y_t^*$ erhält man das gleichgewichtige Preisniveau und den gleichgewichtigen Output als

$$p_t = E_{t-1}(m_t) + a(m_t - E_{t-1}(m_t)) - y_t^* + a(v_t - u_t) \quad (4.14)$$

$$y_t = y_t^* + (1-a)(m_t - E_{t-1}(m_t)) + (1-a)v_t + au_t \quad (4.15)$$

wobei $a = (1/1 + \beta)$ ist. Die Gleichungen (4.14) und (4.15) zeigen einige wichtige Punkte des Modells: Die Geldmengenerhöhung, die antizipierbar ist, also $m_t = E_{t-1}(m_t)$, wirkt nur auf das Preisniveau. Der Output bleibt auf seinem natürlichen Niveau. Dagegen hat die Komponente, die nicht antizipiert ist, $m_t - E_{t-1}(m_t)$, reale Effekte: sie erhöht (über den Realkasseneffekt) die aggregierte Nachfrage. Die Erklärung dafür beruht auf Lucas' Ableitung der Angebotsfunktion: Die rational agierenden Akteure gehen angesichts des Problems der Signalentnahme davon aus, dass diese Nachfrageerhöhung zum Teil der Verbesserung ihrer relativen Preise geschuldet ist. Dementsprechend erhöht sich das Angebot (an Gütern und Arbeit).

Um die Gleichgewichtslösung zu vervollständigen, setzen wir $m_t - E_{t-1}(m_t) = \varepsilon_t$ in die Gleichungen (4.14) und (4.15) ein. Dann folgt:

$$p_t = (b_0 + b_1 y_{t-1}) - y_t^* + a\varepsilon_t + a(v_t - u_t) \quad (4.16)$$

$$y_t = y_t^* + (1-a)\varepsilon_t + (1-a)v_t + au_t \quad (4.17)$$

Aus beiden Gleichungen ist ersichtlich, dass sich die Ökonomie entlang des Pfades des natürlichen Outputs bewegt. Die Faktoren, die das tatsächliche vom natürlichen Outputniveau abweichen lassen, sind die Zufallsvariable und der Prognosefehler. Sofern keine zufälligen Schocks und Prognosefehler (dies bedeutet keine monetären Schocks) auftreten – sie spielen sowieso langfristig keine Rolle für die Realökonomie, da sie sich langfristig ausgleichen werden –, bestehen keine (konjunkturellen) Schwankungen. Wir diskutieren nun, welche wirtschaftspolitischen Konsequenzen die neuklassische Ökonomie hieraus ableitet.

4-3-2 Die Hypothese der Politik-Ineffektivität

In der traditionellen wirtschaftspolitischen Debatte zwischen keynesianischer – gemeint der Einkommenskeynesianismus – und monetaristischer Ökonomie herrscht der Dezinismus, dass die Ökonomie wirtschaftspolitisch so nah wie möglich am „Vollbeschäftigungspfad“ gehalten bzw. dahin gesteuert werden kann. Hinter dieser Gemeinsamkeit verbirgt allerdings ein entscheidender Unterschied im wirtschaftspolitischen Vorgehen. In diesem Kontext bildet die Beschaffenheit der optimalen Rückkopplungsregeln für den Einsatz der geldpolitischen Instrumente den zentralen praktischen Streitpunkt zwischen den beiden Schulen.

Die keynesianische Ökonomie fordert – indem oder weil sie eine fundamentale Instabilität der Marktökonomie postuliert (sprich: „market failure“) – eine „aktive“ Wirtschaftspolitik. Damit verbunden ist eine Rückkopplung zwischen laufenden und früheren Beobachtungen des Zustandes der Ökonomie und der zukünftigen Festlegung der geldpolitischen Instrumente. (Im Zusammenhang mit dem dargestellten Modell bedeutet dies eine Manipulierung der Parameter b_0 und b_1). Für diese Position wurden die Metaphern als „leaning against the wind“ oder „look at everything, respond to everything“ geprägt. Im Gegensatz dazu treten die Monetaristen aufgrund der Annahme der Markteffizienz dafür ein, dass die Geldpolitik als Regel ohne Rückkopplung festgelegt werden soll. Die Geldpolitik soll, ohne Rücksicht darauf, welche allgemeine Konjunkturlage gerade herrscht, ein konstantes Wachstum der Geldmenge anstreben. Dies ist die k -Prozentregel – wobei der Prozentsatz k sich aus dem Potentialwachstum des Outputs und der angestrebten Inflationsrate ergibt.

Sargent/Wallace (1975: 201) haben herausgestellt, dass die logische Struktur jeder keynesianischen Theorie der Wirtschaftspolitik auf folgenden Annahmen aufgebaut seien:⁵⁷ (i) Die Modellstruktur ist durch eine extensive Simultaneität gekennzeichnet, so dass politische Schocks, die eine Variable berühren, auch einen Einfluss auf die meisten anderen ausüben; (ii) da in diesem System Verzögerungen („lags“) existieren, werden die Wirkungen politischer Schocks auf endogene Variablen im Zeitverlauf verteilt. Sie sind seriell korreliert und mithin prognostizierbar; (iii) die Struktur dieser Verzögerungen bleibt im Zeitverlauf konstant, ungeachtet der Änderung im geldpolitischen Verhalten. Daraus folgt, dass die monetäre Autorität im Allgemein ihre geplanten Parameter für die politi-

⁵⁷ Wenn die neuklassische Ökonomie von der keynesianischen Theorie der Wirtschaftspolitik spricht, meint sie damit die optimale Kontrolltheorie vom Typ Tinbergen-Theil.

schen Instrumente jederzeit dann revidieren sollte, wenn sie neue und überraschende Daten über die Variable erhält, die es zu steuern gilt, wie das BSP oder die Arbeitslosenquote.

Sargent/Wallace bestreiten diese Ansicht aus der Perspektive der rationalen Erwartungen, wonach die Wirtschaftssubjekte alle möglichen Informationen nutzen und diese mittels des wahren Modells – die Quantitätstheorie für die neuklassische Ökonomie – verarbeiten. Sie beobachten die Geldpolitik der Zentralbank und reagieren auch auf die geldpolitischen Aktionen. Die rationale Erwartungshypothese erbringt somit andere Ergebnisse als die keynesianische Rückkopplungspolitik unterstellt. Erstens wird eine Geldpolitik, die irgendeinem Muster folgt – dies gilt auch für die Rückkopplungsregeln –, von den Wirtschaftssubjekten antizipiert und bleibt daher neutral; zweitens wird die Invarianzannahme, die für die Rückkopplungsregeln erforderlich ist, nicht zutreffen, weil die wirtschaftspolitischen Maßnahmen die Verhaltensparameter verändern. Aus dieser Perspektive haben Sargent/Wallace die Politik-Ineffektivität deduziert.

Die Hypothese der Politik-Ineffektivität kann formal in folgender Weise dargestellt werden. Wie oben beschrieben, richtet sich die Stabilisierungspolitik darauf, das tatsächliche Outputniveau y_t nah am natürlichen Niveau zu halten, anders ausgedrückt: $(y_t - y_t^*)$ zu minimieren. Technisch handelt es sich um die Minimierung der mittleren quadratischen Abweichung („mean squared error“):⁵⁸

$$MSE = E(y_t - y_t^*)^2$$

Die Hypothese der Politik-Ineffektivität besagt, dass Geldpolitik ineffektiv in ihrer Wirkung auf diesen MSE ist, in dem Sinne, dass der MSE nicht von der geldpolitischen Wahl der Parameters b_0 und b_1 in Gleichung (4.13) (die beiden Parameter bildet die systematische Komponente des geldpolitischen Verhaltens) abhängig ist. Man kann dies leicht sehen. Aus der Gleichung (4.17) geht hervor

$$(y_t - y_t^*)^2 = (1-a)^2 \varepsilon_t^2 + ((1-a)v_t + au_t)^2 + 2(1-a)\varepsilon_t \cdot (1-a)v_t \cdot au_t \quad (4.18)$$

⁵⁸ Dazu vgl. Sargent (1994: 356ff.) und McCallum (1989: 222ff.)

Da definitionsgemäß $E(x)^2 = \sigma_x$ (Varianz von x) und die Kovarianz von x und y $Cov(x, y) = 0$ ist (wegen der angenommenen Unabhängigkeit der Störterme), wobei x, y für beliebige Zufallsvariable mit den Mittelwerten von Null und mit endlichen Varianzen stehen, ergibt sich als *MSE*

$$E(y_t - y_t^*) = (1-a)^2 \sigma_\varepsilon^2 + (1-a)^2 \sigma_v^2 + a^2 \sigma_u^2 \quad (4.19)$$

Aus der Gleichung (4.19) folgt unmittelbar die Ineffektivität der Politik: Die beiden Parameter b_0 und b_1 treten nicht in der Bestimmungsgleichung für den *MSE* auf, was impliziert, dass die unterschiedlichen Werte von diesen Parametern keine Wirkung auf den *MSE* haben. Der Verlauf des tatsächlichen Outputniveaus bleibt von der Politikwahl unberührt.⁵⁹

„In this system, there is no sense in which the authority has the option to conduct countercyclical policy. To exploit the Phillips curve, it must somehow trick the people. But by virtue of the assumption that expectations are rational, there is no feedback rule that the authority can employ and expect to be able systematically to fool the public. This means that the authority cannot expect to exploit the Phillips curve even for one period.” (Sargent/Wallace 1975: 207)

Diese Feststellung ist jedoch nicht zwingend. Z.B. Fischer (1977) und Taylor (1979) haben mit dem „staggered contracts“-Modell gezeigt, dass trotz und wegen der rationalen Erwartungen die Hypothese der Politik-Ineffektivität nicht stichhaltig ist (dazu Kapitel 6). Es sollte bedacht werden, dass die Schlussfolgerung von Sargent/Wallace auf der Annahme der rationalen Erwartungen in Verbindung mit der Lucas-Angebotsfunktion basiert: Indem die rationalen Erwartungen so verwendet werden, dass die Erwartungen identisch mit den Ergebnissen Modells sind, wobei das Modell die Erwartungen determiniert, ist die Neutralität der systemischen Geldpolitik abgesichert. Des weitern wird mittels der Lucas-Angebotsfunktion postuliert, dass die Ökonomie bei ungestörtem Marktmechanismus, speziell ohne Überraschung im Preisniveau, bei Vollbeschäftigung bzw. bei dem mit Vollbeschäftigung kompatiblen natürlichen Niveau bleibt. Bei Lockerung dieser Annahmen bricht die Hypothese der Politik-Ineffektivität in sich zusammen.

⁵⁹ Eine auffällige Implikation ist, dass die Politik-Ineffektivität ungeachtet der Form der politischen Regeln resultiert. McCallum (1989: 224) pointiert, dass dieses Resultat darauf beruht, dass das Verhalten von den Erwartungen abhängig ist, die ihrerseits von den gesetzten Politikregeln abhängen.

4-3-3 Lucas-Kritik

Die Lucas-Kritik, die nach dem Titel von Lucas (1976) benannt wurde, bezieht sich auf die etablierte Forschungspraxis, die die Konsequenzen alternativer Politikzenarios evaluiert unter der Annahme, dass die Parameter des Modells bei der Veränderung der Politik invariant bleiben. Lucas stellt fest, dass diese Vorgehensweise logisch nicht richtig ist, wenn die Erwartungen rational sind. Die Entscheidungsregeln der Wirtschaftssubjekte ändern sich mit der Veränderung des Umfeldes. Die Invarianzannahme ist damit der Lucas-Kritik ausgesetzt.

Stellen wir die Lucas-Kritik in allgemeiner Form dar. Die Standardtheorie der Wirtschaftspolitik basiert auf einem aggregierten Modell, das wie folgt beschrieben werden kann

$$y_t = F(y_{t-1}, x_t, \theta, \varepsilon_t) \quad (4.20)$$

wobei y_t die Zielvariable, x_t die Politikvariablen, θ die Parameter von Funktion F und ε_t die Residualkomponenten der Gleichung sind. Das ökonometrische Modell schätzt den Parametervektor θ , der erklärt, in welchem Ausmaß y_t in Reaktion auf die Änderungen von x_t verändert wird.

Nach der Diagnose von Lucas hat die traditionelle Makroökonomie die Funktion F und den Parametervektor θ aus den Entscheidungsregeln der Akteure (Nachfrage- und Angebotsfunktion) abgeleitet, und unterstellt, dass diese Entscheidungsregeln theoretisch optimal sind. Sie prognostiziert dann die Performance für die unterschiedlichen Werte von x_t auf Basis von invarianten θ . Für Lucas ist diese Vorgehensweise illegitim: „To assume stability of (F, θ) under alternative policy rules is thus to assume that agents' views about the behavior of shocks to the system are invariant under changes in the true behavior of these shocks. Without this extreme assumption, the kinds of policy simulation called for by the theory of economic policy are meaningless.” (Lucas 1976: 25). Der Grund ist, so Lucas, dass die Entscheidungsregeln der Akteure nicht nur von der bekannten Vergangenheit, sondern auch von den Erwartungen über die zukünftige Umwelt abhängig sind und dass die Ankündigung von Politik deshalb diese Entscheidungsregeln verändert.

Im Gegensatz zum Modell (4.20) soll sich ein alternatives Modell für die Theorie der Wirtschaftspolitik spezifizieren lassen als

$$y_t = F(y_{t-1}, x_t, \theta(\lambda), \varepsilon_t) \quad (4.21)$$

wobei der Parametervektor λ die Politikregeln einschließt. Die Änderungen in x_t verändern λ , sodass für die korrekte Prognose erforderlich ist, die Reaktion von θ auf λ zu schätzen. Dabei beeinflusst die Änderung von λ das Systemverhalten in zweierlei Weise: Erstens durch Veränderung des Zeitreihe-Verhaltens von x_t ; zweitens durch die Modifikation der Verhaltensparameter $\theta(\lambda)$, die den Rest des Systems regelt. Allerdings hängt der Einfluss der politischen Parameter auf das Entscheidungsverhalten von der Art und Weise ab, wie die Politik implementiert wird (Lucas 1976: 40f.): Im Fall der diskretionären Politik wird die Bewegung $\theta(\lambda)$ unsystematisch und damit ökonometrisch nicht prognostizierbar; wenn die Politik dagegen anhand von Regeln, die die Implementierung der Politik völlig verständlich machen, durchgeführt wird, kann die resultierenden Strukturveränderungen auf Basis der Schätzung der vergangenen Daten über $\theta(\lambda)$ prognostiziert werden. Lucas stellt damit fest:

„(G)iven that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.“
(Lucas 1976: 41)

Man kann diesen Umstand mithilfe des im vorherigen Kapitel dargestellten Modells verdeutlichen. Die traditionelle Methode geht statt der Gleichung (4.15) davon aus

$$y_t - y_t^* = \phi_0 + \phi_1 m_t + \phi_2 m_{t-1} + \xi_t \quad (4.22)$$

Nun bestimmt man $E_{t-1} m_t$ aufgrund der Gleichung (4.13) als $b_0 + b_1 m_{t-1}$ und setzt dies in Gleichung (4.15) ein. Man erhält

$$y_t - y_t^* = \beta(m_t - b_0 - b_1 m_{t-1}) + (\beta/1 - \beta)v_t + (1/1 + \beta)u_t \quad (4.23)$$

Der Vergleich zwischen (4.22) und (4.23) liefert die Werte der Parameter der Gleichung (4.22): $\phi_0 = \beta b_0$, $\phi_1 = \beta$, $\phi_2 = \beta b_1$ und $\xi_t = (\beta/1 - \beta)v_t + (1/1 + \beta)u_t$. Die traditionelle Theorie der Wirtschaftspolitik hat auf Basis der Gleichungen (4.12) und (4.22) sowie (4.13) die komparativ-statische Analyse durchgeführt. Aber hier offenbart sich, dass die Parameter in Gleichung (4.22) mit der Änderung der Politikparameter gleichzeitig verändert werden. Damit wird die traditionelle Theorie der Lucas-Kritik ausgesetzt. McCallum (1989: 229) weist darauf hin, dass die Gleichung (4.22), die der traditionellen Makroökonomie zugrunde liegt, zwei Verhaltensfunktion verwechselt: nämlich das Angebotsverhalten, das in Gleichung (4.15) zusammengefasst ist, und das Erwartungsverhalten, das im

$E_{t-1} = b_0 + b_1 m_{t-1}$ ausgedrückt ist. Während die erste Verhaltensfunktion definitionsgemäß von der Politik unberührt bleibt, hängen die Erwartungen von dem aktuellen Politikprozess ab.

Aus (monetär)keynesianischer Perspektive ist die Lucas-Kritik zu akzeptieren, soweit sie unterstreicht, dass eine Regeländerung die Verhaltensparameter in eine eindeutige bestimmbare Richtung verändert, wobei die Änderungsrichtung bei rationalen Erwartungen modellabhängig – keynesianisch oder neoklassisch – ist (Schelkle 2001: 6f.). Die Lucas-Kritik lässt sich so auslegen, dass sich die Wirtschaftspolitik ihre eigenen Voraussetzungen schafft und Wirtschaftspolitik und Marktprozess rekursiv werden. Dieser Umstand impliziert wiederum, dass die Wirtschaftspolitik, speziell: die Zentralbank, keine Autorität, sondern Marktteilnehmerin ist. Die Lucas-Kritik decouvriert damit die Defizite der wirtschaftspolitischen Konzeption des Ziel-Mittel-Ansatzes. So gesehen ist die neoklassische Schlussfolgerung, auf eine „aktive“ Wirtschaftspolitik sei zu verzichten, nicht zwingend. Vielmehr kann man die Lucas-Kritik für die Begründung einer aktiven Wirtschaftspolitik in Anspruch nehmen.

4-3-4 Zeitinkonsistenz und politische Regeln

Wir wenden uns in diesem Abschnitt einem Dauerthema in der geldpolitischen Diskussion zu: „rules versus discretion“. Der Gegenstand der Debatte hat sich mit der Zeit geändert. Die frühere Debatte war durch die Auseinandersetzung mit der keynesianischen Vorstellung geprägt. Die Frage war, ob die Geldpolitik überhaupt die ökonomischen Prozesse zielgenau steuern kann. Die Befürworter einer Regelbindung der Geldpolitik, z.B. Friedmans potentialorientierte Geldmengenregel, haben argumentiert, dass die Steuerungsmöglichkeiten der Geldpolitik durch lange und variable Wirkungszögerungen sehr begrenzt seien.

Seit der Veröffentlichung von Kydland/Prescott (1977), deren Idee von Barro/Gordon (1983a, 1983b) weiter entwickelt wurde, steht im Mittelpunkt der Debatte nicht mehr die Frage nach der aktiven vs. nicht-aktiven Geldpolitik. Vielmehr wurde davon ausgegangen, dass die monetären Autoritäten als „wohlmeinende Diktatoren“ die wirtschaftlichen Entwicklungen jederzeit perfekt steuern können (Bofinger u.a. 1996: 137).⁶⁰ Die Regelbindung der Geldpolitik wurde hier unter dem

⁶⁰ Konkret ausgedrückt, wird unterstellt, dass die Zentralbanken die Inflationsrate direkt kontrollieren können, die über die Phillipskurve im Zusammenhang mit Output bzw. Beschäftigung steht.

Schlagwort der „(dynamischen) Zeitinkonsistenz“ optimaler Geldpolitik neu begründet. Die Zeitinkonsistenz definiert sich durch die Differenz zwischen ex ante und ex post Optimalität, d.h., eine optimale Politik zum Zeitpunkt t ist zeitinkonsistent, wenn Re-Optimierung zum Zeitpunkt $t + n$ eine andere optimale Politik impliziert.

Bevor auf die Zeitinkonsistenzansätze eingegangen wird,⁶¹ ist es angebracht, Regeln und Diskretion begrifflich zu definieren. Die Politik ist diskretionär, wenn sie auf einer „period-by-period“ Basis ohne notwendige Verbindung zwischen den Perioden durchgeführt wird. Man spricht von einem optimalen Plan der Geldpolitik in dem Sinne, dass die Politikträger bei gegebener Situation ihre „beste“ Aktion wählen. Im Gegensatz dazu wird die Politik als regelgebunden bezeichnet, wenn sie in jeder Periode nach einer Regel oder Formel, die nicht allein für die laufende Periode sondern für eine ganze Reihe von Periode geltend ist, implementiert wird. Da diese begriffliche Distinktion nicht zutreffend die Praxis der Zentralbank reflektiert, wird sie in der gegenwärtigen Diskussion dahingehend re-interpretiert, dass systematische Reaktion der Politik eine notwendige Bedingung für die Regel und der Einbezug der Wirkungen auf die Erwartungsbildung die hinreichende Bedingung ist. Für eine diskretionäre Politik bedeutet dies, dass sie die Erwartungen als gegeben und daher als ausbeutbar betrachtet.

Barro/Gordon (1983a, 1983b) haben ein Modell präsentiert, um die Vorteilhaftigkeit der Regelbindung selbst unter optimalen Voraussetzungen für eine diskretionäre Politik zu demonstrieren. Ihr Modell ist aus drei Bestandteilen konstruiert: der sozialen Wohlfahrtsfunktion als Zielfunktion der Zentralbank mit der Phillipskurve als Nebenbedingung sowie der rationalen Erwartungsbildung. Die Leitidee war dabei, dass die optimale Kontrolltheorie, die der diskretionären Politik im oben definierten Sinne zugrunde liegt, im Fall des dynamischen ökonomischen Systems mit rationalen Erwartungen ihre Gültigkeit verliere (Kydländ/Prescott 1977: 620). In diesem Fall entsteht das Phänomen der Zeitinkonsistenz, weil die Entscheidungsgrundlage sich ändert: Sobald das Publikum die Absicht der Zentralbank erkennt, antizipiert es die zukünftige Inflation, so dass das Ergebnis suboptimal wird. Der Umstand, dass die rational agierenden Wirtschaftssubjekte das Politikproblem und dessen Lösung in ihrer Erwartungsbildung vorwegnehmen, weist das Modell als spieltheoretische Problemstruktur auf – ein dynamisches Spiel zwischen Zentralbank und Publikum. An dieser Stelle ist zu unterstreichen, dass das Barro-Gordon-Modell, oder allgemeiner: die Zeitinkonsistenzansätze, den analytischen Fokus nicht auf die Marktprozesse und Wirkungen der Geldpolitik

⁶¹ Als Referenz diskutieren wir in diesem Abschnitt das Modell von Barro/Gordon (1983a, 1983b).

setzt. Im Zentrum der Analyse stehen vielmehr die geldpolitischen Entscheidungen – ein Feld, in dem die Spieltheorie ihre Anwendung findet (Spahn 1999).

Die soziale Wohlfahrtsfunktion lautet wie folgt:

$$L_t = a(u_t - k u_t^n)^2 + b\pi^2 \quad \text{mit } a, b > 0, \quad 0 < k < 1 \quad (4.24)$$

Die Wohlfahrtsfunktion ist als Kostenfunktion formuliert, wobei soziale Kosten zwei Ursachen haben können: die Abweichung der tatsächlichen Arbeitsquote von der Zielarbeitslosenquote ($u_t - k u_t^n$) und die Inflationsrate π_t , wobei als Zielinflationsrate von Null unterstellt wird. Die positiven Abweichungen werden ebenso wie die negativen als wohlfahrtsschädlich angenommen. Ein bedeutendes Merkmal ist die Einführung des Faktors $0 < k < 1$, sodass die angestrebte Arbeitslosenquote niedriger als die natürliche ist, also $k u_t^n < u_t^n$. Damit ist unterstellt, dass es in jeder Ökonomie allokativen Verzerrungen gibt und dies die Zentralbank dazu veranlasst, mittels Inflation Beschäftigungseffekte erzielen zu wollen. Das heißt, dass die Zentralbank einen systematischen Inflationsbias hat. Für $k = 1$, dann entfällt der Grund für den Inflationsbias der diskretionären Politik. Die Gewichtungsfaktoren a und b drücken die Präferenz der Zentralbank aus: $a > b$ bedeutet einen hohen Stellenwert des Beschäftigungsziels relativ zum Inflationsziel. Umgekehrt steht $b > a$ für eine hohe Stabilitätspräferenz der Zentralbank.

Als Nebenbedingung wird die Phillipskurve unterstellt, deren Begründung sich in der Lucas-Angebotsfunktion findet.

$$u_t = u_t^n - \gamma(\pi_t - \pi_t^e) \quad \text{mit } \gamma > 0 \quad (4.25)$$

Die optimale Inflationsrate, die von der Zentralbank angestrebt wird, lässt sich errechnen, wenn man die Zielfunktion L_t unter der Nebenbedingung (4.25) nach der Inflationsrate π_t ableitet und die Bedingung erster Ordnung gleich null setzt.⁶² Als Lösung gilt:

$$\pi_t^* = \frac{a\gamma(1-k)}{b+a\gamma^2} u_t^n + \frac{a\gamma^2}{b+a\gamma^2} \pi^e \quad (4.26)$$

⁶² Die Bedingung zweiter Ordnung für die Kostenminimierung ist erfüllt.

Aus der Gleichung (4.26) erhellt, dass die Höhe der optimalen Inflationsrate von der erwarteten Inflationsrate bestimmt wird. Und selbst bei der erwarteten Inflationsrate von Null ist die optimale Inflationsrate positiv, weil unter den getroffenen Annahmen der erste Term in (4.26) stets größer als null ist. Dieses Resultat verdankt sich der Annahme $0 < k < 1$ - eine Annahme, welche den Inflationsbias und damit die Zeitinkonsistenz der optimalen Geldpolitik begründet.

Man kann nun die Lösungen unter unterschiedlichen Politiktypen vergleichen. Unter der diskretionären (Geld)Politik betrachtet die Zentralbank die Inflationserwartungen der privaten Akteure π_t^e als gegeben, während sie die tatsächliche Inflationsrate π_t^{disc} wählt, die die Kostenfunktion L_t minimiert. Die unterstellten rationalen Erwartungen implizieren jedoch, dass die privaten Akteure das Kalkül der Zentralbank kennen. Sie errechnen damit ihre erwartete Inflationsrate durch die Lösung des Optimierungsproblems, dem die Zentralbank gegenübersteht, und prognostizieren die von der Zentralbank angestrebte Inflationsrate. Es handelt sich hier um ein nicht-kooperatives Spiel zwischen Zentralbank und privaten Akteuren. Das Gleichgewicht ergibt sich als Nash-Lösung.⁶³ Die Inflationsrate unter der diskretionären Geldpolitik realisiert sich als

$$\pi_t^{disc} = \pi_t^e = a\gamma(1-k)u_t^n \quad (4.27)$$

Da nach den getroffenen Annahmen die Inflationsschocks null sind, belaufen sich die sozialen Kosten auf

$$L_t^{disc} = a(1+ab\gamma^2)[(1-k)u_t^n]^2 \quad (4.28)$$

Wenn sich die Zentralbank auf eine Regel verpflichtet, schreibt die „beste“ Regel das Inflationsziel von Null für alle Perioden vor. Mit der Regelbindung verliert zwar die Zentralbank die Möglichkeit,

⁶³ Ein Nash-Gleichgewicht ist dadurch gekennzeichnet, dass jeder Spieler genau die Strategie wählt, die ungeachtet der gewählten Strategien der anderen Spieler optimal ist. Die Existenz einer Nash-Lösung als Kombination dominanter Strategien setzt voraus, dass jeder Spieler eine dominante Strategie besitzt und die Spieler simultan über ihre Strategie entscheiden. Im oben dargestellten Modell müssen jedoch die privaten Akteure vorher ihre Erwartungen festlegen (sonst käme die von der Zentralbank unterstellte Phillipskurve nicht zustande). In diesem Fall handelt es sich deshalb vielmehr um ein Stackelberg-Gleichgewicht, bei dem die Zentralbank bei gegebenen Erwartungen der privaten Akteure ihre Kostenfunktion minimiert. Aber unter Annahme der rationalen Erwartungen, wonach die privaten Akteure die optimale Lösung der Zentralbank vorwegnehmen, stimmen die Lösungen für das Stackelberg-Gleichgewicht und das Nash-Gleichgewicht überein (Bofinger 1996: 147f.). Dazu siehe auch Snowdon u.a. (1994: 207f.).

die Arbeitslosenquote unter die natürliche zu senken, dafür entstehen jedoch keine sozialen Kosten der Inflation. Die Inflationsrate und die sozialen Kosten unter der regelgebundenen Geldpolitik sind

$$\pi_t^{rule} = 0 \quad (4.29)$$

$$L_t^{rule} = a \left((1-k) u_t^n \right)^2 \quad (4.30)$$

Betrachten wir nun den Fall, in dem die Zentralbank eine Neigung hat, die Regel zu brechen. Wenn die privaten Akteure die Inflationsrate von Null erwarten, dann würde die Zentralbank zur Überraschungsinflation angereizt, um positive Beschäftigungseffekte zu erzielen. In diesem Fall ergibt sich

$$\pi_t^{cheat} = \frac{a\gamma(1-k)}{b+a\gamma^2} u_t^n \quad (4.31)$$

$$L_t^{cheat} = \frac{ab}{b+a\gamma^2} \left((1-k) u_t^n \right)^2 \quad (4.32)$$

Vergleicht man die drei Ergebnisse, dann ist eine eindeutige Hierarchie der Lösungen festzustellen. Der Fall der Überraschungsinflation weist die geringsten sozialen Kosten auf, gefolgt vom Fall der Regelbindung und der diskretionären Politik.

$$L_t^{disc} < L_t^{rule} < L_t^{cheat} \quad (4.33)$$

Allerdings ist darauf hinzuweisen, dass sich die beste Lösung der Überraschungsinflation nicht einstellen wird, weil es um ein nicht-kooperatives Spiel zwischen Zentralbank und privaten Akteuren geht: Die privaten Akteure können also die Zentralbank für eine Überraschungsinflation in Periode t mit höheren Inflationserwartungen in den Folgeperioden „bestrafen“. Die Zentralbank wird somit ihre Glaubwürdigkeit verlieren, deren Wiederherstellung eine konsequente Disinflationpolitik verlangt, die mit hohen Kosten der Arbeitslosigkeit verbunden sein wird. Aufgrund dieser Ergebnisse folgert das Barro-Gordon-Modell, dass die Regelbindung selbst dann die überlegene Lösung sei, wenn optimale Voraussetzungen für die diskretionäre Politik gegeben sind.

Die Schlussfolgerungen des Barro-Gordon-Modells lauten, dass erstens eine optimale Politik, worunter eine aus der optimalen Kontrolltheorie gewonnene politische Planung verstanden wird, unter den rationalen Erwartungen das Zeitinkonsistenzproblem aufweist und daher in Suboptimalität landet, denn jede wirtschaftspolitische Maßnahme hat einen Erwartungen generierenden und ver-

ändernden Effekt; um dies Zeitinkonsistenzproblem zu lösen, ist zweitens eine (starre) Regelbindung der Geldpolitik notwendig, welche als Fixpunkt gegenwärtiger und zukünftiger Politikpläne fungiert. Das Modell der Zeitinkonsistenz steckt den Rahmen für die Themen ab, die derzeit in der geldpolitischen Diskussion dominieren, nämlich: die Förderung der Glaubwürdigkeit der Geldpolitik (man spricht hier von „Stabilitätsbonus“) und die Gewährleistung der Unabhängigkeit der Zentralbank etc. – Themen, die spieltheoretisch orientiert sind.

Bedenkt man jedoch, dass die Argumente von Zeitinkonsistenzansätzen auf sehr strikten Annahmen beruhen, kann man sagen, dass ihre Anwendung sehr begrenzt bleibt. Besonders fragwürdig ist die Annahme eines k -Wertes kleiner als eins. (vgl. Bofinger u.a. 1996: 150ff.). Seine Bedeutung im Barro-Gordon-Modell besteht darin, dass er die Zeitinkonsistenz der Zentralbank und damit ihren Inflationsbias begründet. Ist die Zentralbank tatsächlich inflationsgeneigt, und stellt die starre Regelbindung die einzige Lösung für die Vermeidung des Inflationsbias dar?

Da keine Zentralbank in der Praxis irgendeiner starren Regel folgt, wurden in der Literatur, die das Zeitinkonsistenzproblem zum Thema hat, andere Lösungen zur Vermeidung des Inflationsbias vorgestellt. Zu nennen seien (vgl. Blinder 1998: 43ff.): das Reputationsproblem der Zentralbanker, das die Glaubwürdigkeit der Geldpolitik verstärkt; optimale Principal-Agent-Kontrakte, die die Zentralbank zur Stabilitätsorientierung anregen. Dabei übernimmt die Regierung die Rolle des Principal und die Zentralbanker die Rolle des Agent; Benennung der konservativen Zentralbanker, deren Inflationsaversion höher als die des repräsentativen Agenten ist (in unserem Zusammenhang bedeutet dies höheren Wert von b als im Modell). Des Weiteren wurden institutionelle Verankerungen der Stabilitätskultur diskutiert: Die (exogene) Festlegung der Preisstabilität als Primärziel der Geldpolitik und die Gewährleistung der Unabhängigkeit der Zentralbank, um geldpolitische Entscheidungen vor einem langfristigen Planungshorizont zu treffen.

Für die Zentralbanker sind jedoch die Annahmen, die dem Barro-Gordon-Modell zugrunde liegen, befremdlich. Blinder (1998: 40ff.) stellt hier eine Lücke zwischen „science“ und „art“ der Geldpolitik fest. Erstens ist die Zentralbank nicht in der Lage, Inflation und Beschäftigung perfekt zu kontrollieren, obwohl das Barro-Gordon-Modell mit dieser Annahme beabsichtigt hat, selbst unter den Idealbedingungen für die diskretionäre Politik die Vorteilhaftigkeit der Regelbindung zu belegen. Zweitens ist der kurzfristige Zinssatz die Steuerungsgröße der Geldpolitik und nicht die monetären Aggregate, wie immer sie definiert werden. Die monetären Aggregate können nicht kontrolliert werden, was eben ihre Endogenität offenbart. Dritten ist der systematische Inflationsbias der Zent-

ralbank ein nicht-existierendes Problem in der Realität.⁶⁴ Die disinflationäre Geschichte seit 1980 zeigt vielmehr das restriktive Verhalten der Zentralbanker, obwohl die Inflation niedrig war und die Arbeitslosigkeit auf hohem Niveau gehalten wurde. Blinder (1998: 43) zufolge zielen die Zentralbanker nicht auf ku_t^n , einfach weil dies inflatorisch wäre. Zudem weist Blinder (1998: 41) darauf hin, dass in der Praxis die Zentralbanker das Problem des Inflationbias nicht durch Adoption einer Regel oder institutionelle Änderungen lösen, sondern aufgrund ihrer aufgeklärten Diskretion: „As in the Nike commercial, they just did it“. “In practice, this ‘just do it’ approach (...) eventually boils down to the fact that the central bank will systematically avoid exploiting temporarily given inflation expectations for the sake of short-lived output gains” (Issing 2001: 36).

Das Barro-Gordon-Modell modelliert das Zeitinkonsistenzproblem so, dass sich ausschließlich die Zentralbank kurzsichtig und irrational verhält. Das Zeitinkonsistenzproblem entsteht, weil ein zentraler Charakter der Zentralbank wegdefiniert ist, sodass der Defekt entsteht. Spahn (1999a: 384) weist darauf hin, dass das Beharren auf dem nichtexistierenden Problem im Barro-Gordon-Modell durch den Fokus auf die falsche Seite der Phillipskurve bedingt ist: Das Barro-Gordon-Modell hat die Überschussnachfrage als Hauptthema behandelt, die jedoch seit den 1970er Jahren ihre Relevanz für die industriellen Länder verloren hat. Es ist eine Ironie, dass die Zeitinkonsistenzansätze auf die Ausbeutung eines Trade-off angewiesen bleiben, obwohl sie in einer Tradition der Kritik an diesem Trade-off stehen.

⁶⁴ “In my view, the academic literature has focused on either the wrong problem or nonproblem and has proposed a variety of solutions (...) that makes little sense in the real world” (Blinder 1998: 50).