NAKE

Nieuws

volume 11, number 1

March 1999

Netwerk Algemene en Kwantitatieve Economie

Netherlands Network of Economics

NAKE Secretariat

Ms. Janna Mesker
Faculty of Economics,
University of Groningen
P.O. Box 800
9700 AV Groningen
Phone: 050-363 6664 / Fax: 050-363 7207.
E-mail: nake@eco.rug.nl

Home Page: http://www.eco.rug.nl/nake

PREFACE

It has been a while since the last issue of *NAKE Nieuws* saw the light of day. This does not mean, however, that nothing has happened on the NAKE front. For one, there have been some drastic changes in the division of tasks in the Board of Management. Gerard van der Laan was succeeded by Joan Muysken as Chairman, Simon Kuipers is the new Secretary, and Theo van de Klundert has taken over the portfolio of Treasurer from our new Chairman. The Board of Managemant will have a tough job ahead of itself in the coming months. Recruitment for the new NAKE Director will commence in the coming weeks, and potential candidates should pay attention to the job ads in the national newspapers and read the electronic *NAKE Bulletin*. NAKE Fellows who know of good candidates should make this known either to me (b.j.heijdra@eco.rug.nl) or to the new Chairman, Joan Muysken (j.muysken@algec.unimaas.nl).

In addition to reshuffling its tasks, the Board of Management has also looked at the issue of inactive NAKE Fellows. A NAKE Fellow is expected to contribute actively to the NAKE activities, e.g. teach a Utrecht course, write reports on NAKE Day thesis proposals, etcetera. As a result of a stock-taking exercise which was conducted during the second half of 1998, there have been some changes in the list of NAKE Fellows. Some formerly active fellows had over the years drifted away from the core fields covered by NAKE and felt that they could no longer contribute to it. The up-to-date list of (by definition) active NAKE Fellows can be found on the NAKE homepage.

In this issue of *NAKE Nieuws* you find (the remaining) two reports on the lectures of the June 1998 workshop which was held in Wageningen. **Gijsbert van Lomwel** (KUB) has written a very good and concise report on the econometrics lectures by Tony Lancaster on "Panel data analyses." Gijsbert has a lot of NAKE brownie points anyway, because in addition to writing the winning Lancaster-interpretation he and **Charles Bos** (EUR) also were excellent local organizers for the December 1998 workshop in beautiful Rotterdam. Wicked voices would argue that there is a quid pro quo here but that is absolute false: the best econometrics report was not even selected by me!

The second report is by **Thijs Knaap** (RUG) who writes very lucidly on the Wageningen lectures by Michael Woodford on "Analysis of monetary policy rules." Thijs is a "repeat offender" in the sense that his report on Ricardo Caballero (June 1996) was also deemed to be the best in its class (see *NAKE Nieuws*, August 1997, **9**(2), 6-13). So either Thijs has bribed me (no!) or he is just a fine young macroeconomist (yes!).

COURSE QUESTIONNAIRE 1999-2000 UTRECHT PROGRAMME

In the middle section of this *NAKE Nieuws* you find a removable course questionnaire plus a listing of all currently existing courses that are offered by the NAKE Fellows. The outlines of (almost all) these courses can be found on the NAKE homepage. Also on the questionnaire you find a listing of proposed courses (bearing a course number preceeded by an “E” for **e**xtra). For some of these courses not much more than the teacher and course title are available at this stage.

Between 16 and 24 courses can be scheduled for inclusion in the Utrecht teaching programme for 1999-2000. I can only design a well-balanced and attractive programme if the potential clientele reveals its (collective) preferences. For this reason I would like to ask all (potential) course participants to fill out the questionnaire and return it to the NAKE secretariat **before May 1, 1999**. Even if you are not sure that you will actually enrol when time comes around, please let us know what you want.

Of course, suggestions for new (i.e. non-existing and currently unplanned) courses are also welcome. We will try to find an appropriate NAKE Fellow to develop and teach the course of your liking. The provisional teaching programme for the academic year 1999-2000 will be announced some time in June 1999 on the NAKE Home Page.

Ben Heijdra

Tony Lancaster
Panel Data Analyses

report by Gijsbert van Lomwel *

November 19, 1998

1 Introduction

The analysis of panel data has become widespread in econometrics. Panel data in general are data sets in which there are multiple observations on the same agents. For instance, one could think of observing multiple unemployment spells of the same unemployed individual. The standard methods to analyze these data can be summarized by three characteristics.

1. Emphasis in existing literature (eg. Hsiao (1985)) on simple estimation, like Least Squares;
2. Frequentist inference;
3. Random effects.

The aim of these lectures by Tony Lancaster is to show how new methods can give more insight into panel data models. These new methods can be characterized, as opposed to the existing methods above, by

1. Computer Intensive, leading to exact[1] inference;
2. Bayesian inference;
3. Fixed effects.

There are several types of panel data models. What follows first is an overview of some of these models. Then the conventional, maximum likelihood based, methods and the new methods as proposed by Tony Lancaster will be discussed. These new methods are demonstrated by an application

*Department of Economics, Tilburg University, P.O.box 90153, 5000 LE Tilburg and CentER

[1]Exact from a Bayesian point of view, i.e. the exact posterior distribution of the parameters

to optimal job search models. This is where the lectures actually ended. This report will end by giving a brief overview (based on the reader) of the last two topics, the incidental parameters problem and its solution, orthogonal parameters. Incidental parameters are also known as nuisance parameters. They arise in panel data models as parameters covering individual specific, time-invariant effects. As opposed to structural parameters, incidental parameters only appear in the probability law of a finite number of random variables, while structural parameters appear in the law of every random variable. As we will see below, incidental parameters can cause maximum likelihood estimates of the structural parameters to be inconsistent. One way of solving this is the orthogonal parameter approach. This trick boils down to diagonalizing the information matrix, by finding a suitable reparametrization of the incidental parameters.

2 panel data models

In all models below, N denotes the number of agents and T denotes the number of observations on each agent.

2.1 Linear Panel

$$y_{it} = v_i + \beta x_{it} + \epsilon_{it} \tag{1}$$

$$\epsilon_{it} \sim i.i.d.\ N(0, \sigma^2) \tag{2}$$

This is the easiest model. In this model, β can be estimated consistently by first differencing the model, thereby loosing the so called fixed effects v_i. However, the maximum likelihood estimator for σ^2 is inconsistent, but this can be easily adjusted for (refer the example by Neyman & Scott (1948), the first occurrence of fixed effects).

2.2 Dynamic Linear Model

$$y_{it} = v_i + \rho y_{it-1} + \beta x_{it} + \epsilon_{it} \tag{3}$$

$$\epsilon_{it} \sim i.i.d.\ N(0, \sigma^2) \tag{4}$$

This model arises for instance in cross-country growth models, addressing the question of growth rate convergence. The fixed effect v_i can then be interpreted as the country specific time invariant factors affecting the level of GDP. In this model there is persistence through v_i and partially through y_{it-1}. The problem is that maximum likelihood estimates for both ρ and σ^2 are inconsistent, because the persistent error terms are correlated with the lagged dependent variable through v_i. First differencing does not help in this case. The standard method to solve these inconsistencies is to take an instrumental variable approach. We will return to another method to solve this problem, the orthogonal parameter approach.

2.3 Panel Count Data

$$y_{it} \sim Poisson(v_i e^{\beta x_{it}}) \tag{5}$$

One could think for example as y being annual patent filings for firm i (Hausman, Hall & Griliches (1984)). This model looks easy, but is not. It suffers from the incidental parameters problem, leading to inconsistent estimates. Again a way to solve this is the orthogonal parameter approach.

2.4 Panel Probit Logit

$$y_{it} \sim Bernoulli(F(v_i + \beta x_{it})) \tag{6}$$

with F some known distribution function. This is equivalent to $Pr(y_{it} = 1|v_i, \beta, x_{i1} \cdots x_{iT}) = F(v_i + \beta x_{it})$. For F equal to the normal distribution function, we get the probit model, and when F is logistic we get the logit model. This model is a fundamental one, because a lot of models can be reduced to this one.

If the v_i are treated as random effects, one can get consistent estimates of β by integrating v_i out. However, if the v_i are treated as fixed effects, maximum likelihood in general gives inconsistent estimates. If F is logistic, there is a solution. In this case there is a sufficient statistic for v_i when β is known, $S_i = \sum_{t=1}^{T} y_{it}$. Then the likelihood of the data $y_{i1} \cdots y_{iT}$ conditional on S_i does not depend on v_i. Other distributions F are not known to have a sufficient statistic for v_i and this problem is still unsolved.

2.5 Dynamic Panel Probit Logit

A dynamic version of the model in the previous subsection is

$$y_{it}|y_{it-1} \sim Bernoulli(F(v_i + \beta x_{it} + \gamma y_{it-1})) \tag{7}$$

with F some known distribution function. This model is still unsolved.

2.6 Panel Duration data

Let

$$y_{it} \sim Exponential(v_i e^{\beta x_{it}}) \tag{8}$$

so that y_{it} has hazard $v_i e^{\beta x_{it}}$. For panel duration models of this, proportional hazards, type, consistent estimates are available under some mild regularity conditions.

2.7 Dynamic Panel Duration data

One could imagine that having experienced a spell of unemployment might have a negative effect on the chance of leaving a second spell of unemployment. This scarring effect is captured by the parameter γ in the following dynamic panel duration model.

$$y_{it}|y_{it-1} \sim Exponential(v_i y_{it-1}^{\gamma} e^{\beta x_{it}}) \tag{9}$$

For this model maximum likelihood gives inconsistent estimates.

2.8 Multiple Discrete State Process

Multiple discrete state processes are an extension of the probit / logit model. In this case the outcome of one process affects the outcome of another process. One might think of the amount of tissues you buy depending on the amount you bought last period, but also on whether or not you had a baby in the previous period;

$$y_{1it}|y_{1it-1}, y_{2it-1} \sim Bernoulli(F(v_i + \beta x_{it} + \gamma y_{1it-1} + \rho y_{2it-1})) \tag{10}$$

As one can imagine, this complicates the analysis even further, compared to the standard probit or logit model.

3 Methods of Analysis

One can take different approaches to attack the consistency / incidental parameters problems, described above. In this lecture these methods are reviewed. Three approaches will be distinguished:

1. Likelihood methods
2. Bayesian methods
3. Sampling Bayes and Data Augmentation

3.1 Likelihood methods

Let $L(\theta), \theta \in \Theta$ denote the log-likelihood. The data having occurred, we can think of the likelihood as the probability of the data conditional on unknowns called parameters. We take the logarithm of this function for mathematical convenience. The derivatives of this log-likelihood function to θ play a key role in the analysis of the maximum likelihood estimator. The score is defined as the first derivative to θ , $\frac{\partial L}{\partial \theta}$. This score has the following fundamental property:

$$E_{\theta=\theta_0}\Big[\frac{\partial L}{\partial \theta}\Big] = 0 \tag{11}$$

Suppose you have a data generating process, described by a parameter θ. The property above means that when you have different realizations of this process with $\theta = \theta_0$, i.e. different data sets, and for each data set you draw the log-likelihood, then on average the slope of the log-likelihood function evaluated at the true parameter θ_0 equals 0. The negative Hessian is defined as

$$-\frac{\partial^2 L}{\partial \theta^2} \tag{12}$$

Related to this is the Information, defined as

$$-E\left[\frac{\partial^2 L}{\partial \theta^2}\right] = \mathcal{I}(\theta) = E\left[\frac{\partial L}{\partial \theta}\right]^2. \tag{13}$$

The maximum likelihood estimator of θ, $\hat{\theta}$, is the solution of the likelihood equations, the first order condition that the score equals 0;

$$\frac{\partial L}{\partial \hat{\theta}} \equiv 0. \tag{14}$$

This estimator has the following properties

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, \mathcal{I}(\theta_0)^{-1}), n \rightarrow \infty. \tag{15}$$

This means that the quantity $\sqrt{n}(\hat{\theta} - \theta_0)$ has a limit in distribution that is a normally distributed random variable. This is the distribution after repeatedly taking an infinite sample from the data generating process with parameter θ_0. Because the mean of this distribution equals 0, the estimator is consistent. It can also be proven that this estimator is asymptotic efficient (see for example Davidson & MacKinnon (1993)).

As an example of how maximum likelihood works, consider a random sample drawn from an $Exponential(ve^{x_i\beta})$ distribution. For instance, this sample could be observations on durations t. Then the contribution to the likelihood of observation i equals

$$l_i = ve^{x_i\beta}e^{-ve^{x_i\beta}t_i}. \tag{16}$$

Taking logs gives the log-likelihood contribution of observation i,

$$L_i = \log v + x_i\beta - ve^{x_i\beta}t_i. \tag{17}$$

We also have

$$\frac{\partial L_i}{\partial v} = \frac{1}{v} - e^{x_i\beta}t_i. \tag{18}$$

Combined with the fact that the expectation of t_i equals $\frac{1}{ve^{x_i\beta}}$, we get the general property stated above that in repeated sampling of t with x_i, β and v fixed, you get on average $\frac{\partial L_i}{\partial v} = 0$, or

$$E\left[\frac{\partial L_i}{\partial v}\right] = 0. \tag{19}$$

To get the maximum likelihood estimate for v, we construct the likelihood equations and solve for v.

$$\frac{\partial L}{\partial v} = \frac{n}{v} - \sum_i e^{x_i\beta} t_i = 0, \tag{20}$$

$$\hat{v} = \frac{n}{\sum e^{x_i\beta} t_i}. \tag{21}$$

We can now lower the dimension of our maximization problem by one, by substituting the above maximum likelihood estimate of v (which is a function of β) into the log-likelihood function[2]. Solving the likelihood equations, which now only depend on β, gives the maximum likelihood estimate for β.

3.2 Bayesian methods

We now consider the same example from a Bayesian perspective. Again we have n independent observations on n individuals with covariates x_i and parameter pair (v, β). Because these observations are drawn from an $Exponential(ve^{x_i\beta})$ distribution, the likelihood of this data set is

$$l = v^n e^{\sum_i x_i\beta} e^{-v\sum_i e^{x_i\beta} t_i}. \tag{22}$$

Central in doing Bayesian inference is Bayes' rule:

$$P(\theta|data) = P(data|\theta) \cdot P(\theta)/P(data) \tag{23}$$

in which θ denotes the vector of parameters to be estimated. In Bayesian jargon, $P(\theta|data)$ is called the *posterior*, $P(\theta)$ the *prior* and $P(data|\theta)$ the classically familiar *likelihood*. $P(data)$ does not depend on the parameters and just serves as a scaling factor to let the posterior integrate to 1. The posterior is a mixture of the prior believes about the parameters and the "current information", ie. the data. In this way one can continuously accumulate information about the parameters, because the posterior can serve as the prior as soon as new information becomes available. Note that the specification of the prior believes about the parameters is a subjective matter, which should be done without taking notice of the data. Objective Bayesians propose to take uninformative priors, stating that the parameters could be anywhere in the parameter space, but this has not ended the discussion between Bayesians and frequentists.

Let's proceed by taking the following prior, denoted by $\pi(v, \beta) \propto \frac{1}{v}$. Then the posterior equals

$$P(v, \beta|data) \propto v^{n-1} e^{\sum_i x_i\beta} e^{-v\sum_i e^{x_i\beta} t_i}. \tag{24}$$

[2] This procedure is called concentrating the likelihood, or profiling the likelihood in statistics

If we are interested in the parameter β, we can integrate out v from the above expression:

$$P(\beta|data) \propto e^{\sum_i x_i\beta} \int v^{n-1} e^{-v\phi} dv,\ \phi = \sum_i e^{x_i\beta} t_i. \tag{25}$$

In the 1-dimensional case (just 1 β), you can graph $P(\beta|data)$. However, what to do in case of a higher dimensional posterior? In case of 15 β's it is at least hard to integrate all other 14 β's out in order to graph the posterior distribution of the one β you are interested in. This is where computer intensive methods like sampling and data augmentation come into play.

3.3 Sampling Bayes and Data Augmentation

Suppose you have a posterior $p(\theta_1, \theta_2|data)$, θ_1, θ_2 both parameter vectors, and you are interested in θ_1. In this case, you can proceed by integrating out θ_2;

$$p(\theta_1|data) = \int p(\theta_1, \theta_2|data) d\theta_2. \tag{26}$$

Now suppose you cannot integrate, for instance because θ_2 is of very high dimension. The solution then is to sample from the distribution $p(\theta_1, \theta_2|data)$, using the Gibbs sampler. To do this, first construct the so called *component conditionals*. In the example above these are:

$$p(\theta_1|\theta_2, data)$$
$$p(\theta_2|\theta_1, data)$$

Using these component conditionals, we can construct a *Gibbs sequence* of random variables

$$\theta_2^0, \theta_1^0, \theta_2^1, \theta_1^1, \ldots, \theta_2^k, \theta_1^k. \tag{27}$$

The initial value θ_2^0 is chosen, and the rest of the sequence is obtained iteratively by alternatively sampling from

$$\theta_1^j \sim p(\theta_1|\theta_2^j, data)$$
$$\theta_2^{j+1} \sim p(\theta_2|\theta_1^j, data).$$

Under certain regularity conditions, the sequence θ_1^j, θ_2^j converges to a stationary distribution, which turns out to be the joint distribution $p(\theta_1, \theta_2|data)$. Therefore, for k large enough, each θ_1^k is an observation (correlated with θ_1^{k-1}) of the marginal distribution. One way to obtain a sample from the marginal distribution is to generate m independent Gibbs sequences of length k and using the final observation of each sequence.

Consider for example the bivariate normal distribution with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ. The analytical expression for this distribution is quite difficult. The component conditionals are univariate normal

$$p(y_1|y_2, data) \sim N\Big(\mu_1 + \frac{\rho\sigma_1\sigma_2}{\sigma_2^2}(y_2 - \mu_2), \sigma_1^2(1-\rho^2)\Big)$$
$$p(y_2|y_1, data) \sim N\Big(\mu_2 + \frac{\rho\sigma_1\sigma_2}{\sigma_1^2}(y_1 - \mu_1), \sigma_2^2(1-\rho^2)\Big).$$

These distributions can be easily sampled from in order to construct Gibbs sequences. It is very illustrative to plot this sequence. In the 2-dimensional case the 99% region of a uni- modal distribution is an ellipse. The chain passes through this ellipse, exploring the region where the probability is non-negligible. Suppose now we have sampled m independent observations y_1^k, y_2^k by the way described above. These observations can be used to estimate the marginal densities. The common sense thing to do would be to take a histogram kernel smoother approach. However, this is not the most efficient thing to do. Most efficient is to use the quantities $p(y_1|y_2)$ because these carry more information about $p(y_1)$ than the y_1^k's alone. By definition we have that

$$p(y_1) = \int p(y_1, y_2)dy_2 = \int p(y_1|y_2)p(y_2)dy_2. \tag{28}$$

This is an average of $p(y_1|y_2)$ at any point y_1. We know the quantities $p(y_1|y_2)$ by definition of Gibbs sampling. This leads to the estimator

$$\hat{p}(y_1) = \frac{1}{m}\sum_{i=1}^{m} p(y_1|y_2^i) \tag{29}$$

i.e. substitute all the final observations of the m sequences y_2^j you have sampled in the known component conditional and average over all m sequences. This method is called the Rao-Blackwell method.

In a way, Gibbs sampling boils down to conditionals determining the marginals. As Casella & George (1992) show, the Gibbs sequence of one of the random variables can be seen as a Markov chain, with the transition probabilities determined by the component conditionals. Like above consider the bivariate case Y_1, Y_2, with marginal probabilities $p_{Y_1}(y_1)$ and $p_{Y_2}(y_2)$. Using the laws of conditional probability you get for $p_{Y_1}(y_1)$ (and analogously for $p_{Y_2}(y_2)$)

$$p_{Y_1}(y_1) = \int h(y_1, t)p_{Y_1}(t)dt, \tag{30}$$

where $h(y_1, t)$ is a function of the component conditionals. The above equation defines a *fixed point integral equation*, for which $p_{Y_1}(y_1)$ is the unique solution. Gibbs sampling works because the iteration scheme in the limit equals the above equation, i.e. as $k \to \infty$

$$p_{Y_1^k|Y_1^0}(y_1|y_1^0) \to p_{Y_1}(y_1)$$
$$p_{Y_1^k|Y_1^{k-1}}(y_1|t) \to h(y_1|t).$$

Now consider a simple probit model

$$y \in \{0,1\}, \quad P(y_i = 1|\alpha, \beta, x_i) = \Phi(\alpha + \beta x_i), \tag{31}$$

where $\Phi(\dot{)}$ denotes the cdf. of the standard normal distribution. The likelihood of a sample of n observations equals

$$l(\alpha, \beta) = \prod_{i=1}^{n} \Phi(\alpha + \beta x_i)^{y_i} \cdot \Phi(\alpha + \beta x_i)^{1-y_i}. \tag{32}$$

Taking an uninformative prior, $\pi(\alpha, \beta) \propto 1$, implies that the posterior is proportional to the likelihood. This posterior is a mess, as are to component conditionals, so Gibbs sampling is a bad idea. However, consider the latent regression belonging to the above probit model,

$$y_i^* = \alpha + \beta x_i + \epsilon_i, \quad \epsilon_i \sim i.i.d.\ N(0,1). \tag{33}$$

In the model $y_i = 1$ iff $y_i^* > 0$. Suppose now we sampled from $p(\alpha, \beta, y_1^*, \ldots, y_n^*|data)$. We don't have to know the analytical expression for this posterior, because the component conditionals follow in an intuitive way.

1. $p(\alpha, \beta|y_1^*, \ldots, y_n^*, x_i, data) = p(\alpha, \beta|y_1^*, \ldots, y_n^*, x_i)$ because the data are redundant when $y_1^*, \ldots, y_n^*$ are known. Given the latent regression and the uninformative prior, this conditional is equal to

   $$p(\alpha, \beta|y_1^*, \ldots, y_n^*, x_i) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\sum_i (y_i^* - \alpha - \beta x_i)^2}.$$

   Note that α and β enter quadratically in this expression, ie. they have a multivariate normal distribution. The posterior means of α and β equal their Least Squares estimate, while the posterior variance equals the Least Squares covariance $(X'X)^{-1}$. To sample from this multivariate normal distribution, univariate normal distributions can be used, applying the Choleski decomposition of the covariance matrix.

2. $p(y_1^*, \ldots, y_n^*|\alpha, \beta, data)$. These y_i^*'s are independent. Now consider $p(y_1^*|\alpha, \beta, data)$. This is a left $(y_1 = 1)$ or right $(y_1 = 0)$ truncated normal distribution with appropriate mean $\alpha + \beta x_i$ and variance equal to 1. So to sample from the component conditional $p(y_1^*, \ldots, y_n^*|\alpha, \beta, data)$ you can sample n times from the appropriate truncated normal distributions.

The above described technique is called Data Augmentation. This works because the Gibbs procedure also works when you take the component conditionals block-wise. In other words, the fixed point integral equation also holds if you consider a pair (Y, Z) as one random variable.

4 An Application

The above described technique of sampling of posterior distributions by data augmentation and Gibbs is particularly convenient in models with a simple latent structure (that is to say you wish you would have observed more). The probit model in the example above is a good example of this. Another is the classical optimal job search model. In this model, job offers are assumed to arrive following a Poisson process. Each job offer is a random draw from a conveniently chosen and parameterized job offer distribution. It will be shown that using the simple latent structure, inference is relatively simple (Lancaster (1997a)).

An agent in this model is supposed to maximize the expected present value of future income streams over an infinite horizon. The optimal policy can be expressed by a reservation wage ξ, which satisfies the optimality condition

$$\xi = b + \frac{\lambda}{\rho} \int_{\xi}^{\infty} \bar{F}(w)\, dw, \tag{34}$$

where b is the benefits level, λ is the job offer arrival rate, ρ the discount rate and $\bar{F} = 1 - F$, F being the cumulative wage offer distribution. The wage offer distribution is assumed to be log-normal with parameters μ and σ. This equation has a unique solution. If the wage offered is higher than ξ, the job is accepted and the agent will have this job infinitely long, otherwise the agent rejects and waits for the next offer (sequential search).

In econometric practice, what is observed are n pairs (w, t) of the duration of search t and the accepted wage w. Let θ refer to the parameters μ, σ, ρ and λ. Because t has an exponential $(\lambda \bar{F}(\xi))$ distribution, and $w|t$ a truncated log-normal distribution, it easily follows that the likelihood for a sample of n observations (w_i, t_i) equals

$$l(w, t|\theta) = \lambda^n e^{-\lambda \bar{F}(\xi) \sum_i t_i} \prod_i f(w_i), \quad b < \xi < \min_i w_i. \tag{35}$$

This is a very difficult expression, not only because $\xi(\theta)$ now enters the kernel of the likelihood. This will not be solved by taking appropriate priors. Assume that we take the prior $\pi(\theta) \propto 1$. Then the above likelihood equals the posterior distribution of θ. Data augmentation enables us to sample from this distribution. What we will do is point out this procedure when ξ is treated as a parameter, i.e. there is a reservation wage ξ, but this is not necessarily formed by the optimality condition stated above.

We augment the data by the unobserved variables s and $u_1, \ldots, u_s$. The variable s denotes the number of rejected offers, and $u_1, \ldots, u_s$ denote the values of these offers. We now want to sample from $p(\theta, \xi, s, u|w, t)$. Consider the (block-wise) component conditionals.

1. $p(\theta, \xi|s, u, w, t)$. It follows that under full observability (i.e. we observe all wage offers, accepted or not), the likelihood function may be taken

as

$$
\begin{aligned}
l(s,u,w,t|\theta,\xi) &= \lambda^{n+S} e^{-\lambda T} \prod_{i=1}^{n+S} f(v_i), \\
& u = \max_i\{b, \max_j u_{ji}\} < \xi < \min_i w_i = w, \quad (36)
\end{aligned}
$$

where S equals $\sum_{i=1}^n s_i$, T equals $\sum_{i=1}^n t_i$ and v_i is the i'th observed offer, rejected or not. This likelihood can be easily derived using standard probability theory. For instance, the probability $p(u|s,w,t)$ is the offer distribution truncated on the right at ξ and the distribution of $s|w,t$ is a poisson distribution. Now, if θ and ξ are stochastically independent a priori, then they also are a posteriori. This means that sampling from $p(\theta,\xi|s,u,w,t)$ is easy. The posterior of ξ is equal to the prior distribution restricted to the interval (u,w). The posterior of θ can also be derived straightforwardly.

2. $p(s,u|\theta,\xi,w,t)$. As noted above, $u|s,w,t$ is the offer distribution truncated on the right at ξ and $s|w,t$ is a poisson distribution. Therefore this posterior is a product over n agents of a poisson distribution for s_i and s_i draws from the offer distribution truncated on the right at ξ. Therefore, sampling is easy: first sample s and then sample from this truncated offer distribution.

In case the optimality condition is enforced, ξ disappears as a parameter and ρ is identified. The output of the Gibbs sampler described above can now be used to sample from the posterior in this case, $p(\theta|s,u,w,t)$. However, the point we want to make here is that the latent structure of the job-search model causes data augmentation to be very applicable. Considering the number of rejected offers and their values produces common distributions in an intuitive way.

A striking outcome of the Bayesian analysis of the job-search model using Gibbs sampling and data augmentation is that the parameters a posteriori appear to be far from being normally distributed, as is assumed in the classical maximum likelihood context.

5 Incidental Parameters

The last part of this lecture was meant to be on the incidental parameters problem and the orthogonal parameters approach to this problem. In panel data models with common parameters (say K) to all agents and individual specific fixed effects (say N), direct application of maximum likelihood to all $N+K$ parameters can lead to inconsistent estimates of the common parameters as the number of individual specific parameters goes to infinity.

In case the fixed effects f_i are likelihood orthogonal to the common parameters λ, possibly after reparametrization of the fixed effects, the incidental parameter problem disappears, because separate inference can be made on λ. In this case

$$\frac{\partial^2 \log l}{\partial f_i \partial \lambda} = 0. \tag{37}$$

In many cases, likelihood orthogonality cannot be achieved. In these cases the frequentist approach is to try to find a reparametrization of f_i, f_i^* such that

$$E\left[\frac{\partial^2 \log l}{\partial f_i^* \partial \lambda}\right] = 0. \tag{38}$$

In this case, f_i^* and λ are parameter or informational orthogonal. This is weaker than the first condition, the second achieves on average what the first does identically. The second condition is equivalent to block-diagonality of the information matrix. There are two intuitive reasons why information orthogonality works. First, block diagonality of the information matrix means that asymptotically f_i and λ are independently distributed, so that separate inference can be done. Secondly, if f_i and λ are information orthogonal, it can be showed that maximum likelihood estimates of λ hardly depend on f_i.

The Bayesian approach to incidental parameters would be to integrate them out with respect to an uninformative prior. However, this is no guarantee for consistent estimates. The uniform prior needs to be carefully chosen. The frequentist orthogonal parameters approach can help. It turns out (Lancaster (1997*b*)) that integrating out orthogonalized fixed effects with respect to an uninformative prior solves the incidental parameters problem in a number of econometric models, eg. the dynamic linear panel data model.

References

Casella, G. & George, E. (1992), 'Explaining the Gibbs sampler', *The American Statistician* **46**(3), 167–174.

Davidson, R. & MacKinnon, J. (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Hausman, J., Hall, B. & Griliches, Z. (1984), 'Econometric models for count data with an application to the patent-R&D relationship', *Econometrica* **52**(4), 909–938.

Hsiao, C. (1985), *Analysis of Panel Data*, Cambridge University Press.

Lancaster, T. (1997*a*), 'Exact structural inference in optimal job-search models', *Journal of Business & Economic Statistics* **15**(2), 165–179.

Lancaster, T. (1997*b*), Orthogonal parameters and panel data.

Lancaster, T. (1998), The incidental parameters problem since 1948.

Neyman, J. & Scott, E. (1948), 'Consistent estimation from partially consistent observations', *Econometrica* **16**, 1–32.

Analysis of Monetary Policy Rules
The NAKE lectures of Michael Woodford

Report by Thijs Knaap

June, 1998

1 Introduction

The lectures that professor Woodford taught at the recent Wageningen version of the NAKE workshop were, in essence, about monetary policy rules. But rather than taking a partial approach, they involved the construction of a complete macro model and submerged the crowd of listeners in the rigorous Rotemberg–Woodford methodology (as also featured in Rotemberg and Woodford, 1992, for example).

Monetary policy rules describe the behavior of a central bank's monetary policy, either *ex ante* as a policy announcement or *ex post* as a rationalization of the bank's actions by economists. The current interest in policy rules is from the side of central bankers as well as academia. The central bank needs an explicit conceptual framework because of increased autonomy, and because many of the old guideposts, such as monetary targets and exchange rate goals, have become obsolete. Also, precommitment to a rule may be a better way to implement a policy than haphazardly fixing the rate because of the way in which consumer expectations are affected.

In academics, the *communis opinio* now is that money matters, not so much as the instigator of shocks, but more in terms of the propulsion of exogenous shocks to the economy. Developments in modelling now allow for a micro-founded analysis of this propulsion that was not possible in the past. Those analyses often call for the formulation of consumers' expectations about monetary policy, which are difficult to formulate without a rule.

A well-known example of a monetary policy rule is by Taylor (1993), who gives the desired interest rate as a function of the steady-state interest and inflation rate (r^* and π^*), trend output y^* and the actual realizations of these variables:

$$r_t = r^* + \pi^* + \phi_\pi \left(\pi_t - \pi^*\right) + \phi_y \left(y_t - y_t^*\right). \quad (1)$$

In the United States, a rule like this is thought applicable to the years of Volcker and Greenspan, 1979 till now. Notice that the variable that is 'doctored' is the interest rate, instead of the usually assumed aggregate money supply. This is thought more in line with the actual practice of monetary policy; also, policy aimed at r does not depend on a stable demand for money.

It would be outside the scope of this report to fully replicate professor Woodford's lectures. By leaving out the non-essential parts, we will try to replicate the gist of the story, referring the interested reader to Rotemberg and Woodford (1997, 1998) for the details. In the next section, we look at the micro-founded macro model that serves as the workhorse for our analysis. Section 3 shows how this model naturally spawns a criterion by which to judge the outcome of a policy-rule. The feasibility of different rules is discussed in section 4, and in section 5 we look at several counterfactuals run over the period 1979–95. Section 6 concludes.

2 A macro model

The aim of this section is to show the construction of a simple macro model, which exhibits the interplay between inflation, aggregate production, and the interest rate. The economy is taken to be a continuum of households, indexed on $[0, 1]$. Each household functions like a yeoman farmer, producing one good and trading it with other households. Money is introduced in the model in the Sidrauski-way, by making households like it *per se*. The utility of agent j is given by

$$U_t^j = E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[u\left(c_t^j\right) - v\left(y_t^j\right) + w\left(\frac{m_t^j}{P_t}\right)\right]\right\} \tag{2}$$

where y_t^j is agent j's production and c_t^j is a CES aggregate of the different products consumed by the agent:

$$c_t^j = \left[\int_0^1 c_t^j\left(z\right)^{\frac{\theta-1}{\theta}} \mathrm{d}z\right]^{\frac{\theta}{\theta-1}} \quad \text{where } \theta > 1. \tag{3}$$

It follows that the market for goods is monopolistically competitive. The functions u and w are increasing and convex, and v is increasing and concave. From the utility function we see that working causes displeasure and the possession of money is valued positively. The price index P_t combines the prices of goods, P_t^j, in line with the consumption index in formula (3).

It is assumed that agents have at their disposal the use of risk-free bonds that pay real interest r_t. The Fisher parity condition implies that nominal interest i_t satisfies

$$1 + i_t = \left(1 + r\right)\frac{P_{t+1}}{P_t}. \tag{4}$$

Thus, agents can shift wealth through time in the form of both money (m_t^j) and bonds (b_t^j). The budget condition that goes with the problem is

$$b_{t+1}^j + \frac{m_t^j}{P_t} = \left(1 + r\right) b_t^j + \frac{m_{t-1}^j}{P_t} + y_t^j - c_t^j, \tag{5}$$

keeping in mind that at time t, the decision is made how much money to hold now and how many bonds to keep until the next period.

Because the agents are both producers and consumers, maximizing their utility gives the equilibrium of the entire economy. The problem is solved by analyzing three sub-problems:

1. given total expenditure, allocate across the different goods.
2. given the revenues from supplying the goods, find an intertemporal plan for $\{c_t^j\}$ and $\{m_t^j\}$.
3. find the optimal pricing (or supply-) decision each period.

Problem *1* is the standard monopolistic competition problem, and renders

$$c_t^j(z) = c_t^j \left(\frac{P_t(z)}{P_t}\right)^{-\theta} \tag{6}$$

Problem *2* is solved by substituting (5) in (2) and differentiating with respect to m_t^j and b_t^j. This gives two efficiency conditions. First, the usual Euler smoothing equation

$$u'\left(c_j^t\right) = \beta(1+r_t) E_t\left[u'\left(c_{t+1}^j\right)\right] \tag{7}$$

holds. Secondly, deferring consumption one period and using the freed-up money to invest in bonds should be utility-neutral, so that[1]

$$\frac{w'(M_t^j/P_t)}{u'(c_t^j)} = 1 - \frac{1}{1+i_t} = \frac{i_t}{1+i_t}. \tag{8}$$

2.1 Frictionless equilibrium

To solve problem *3*, we need to assume a type of price-setting behavior. Assume first that prices are perfectly flexible. Agents solve the problem

$$\max_{P_t^j} \lambda_t^j \left(P_t^j \mathrm{Y}_t \left(\frac{P_t^j}{P_t}\right)^{-\theta}\right) - v\left(\mathrm{Y}_t \left(\frac{P_t^j}{P_t}\right)^{-\theta}\right)$$

where λ_t^j is the marginal utility of additional money revenue[2] at time t, and Y_t is and index for total consumption, $\mathrm{Y}_t = \int_0^1 c_t^j \mathrm{d}j$. The f.o.c., taking into account that P_t and Y_t are not affected by changes in P_t^j, is

$$P_t^j = \frac{\theta}{\theta-1} \frac{v'\left(\mathrm{Y}_t\left(\frac{P_t^j}{P_t}\right)^{-\theta}\right)}{\lambda_t^j}.$$

[1] This condition may be found by taking (2)'s first order condition w.r.t. m_t^j, and using the Euler equation (7) and the Fisher parity (4).

[2] There holds that $\lambda_t^j = u'\left(c_t^j\right)/P_t$.

Because in equilibrium, all agents act the same, we can substitute in that $c_t^j = \text{Y}_t$, $m_t^j = M_t$ and $P_t^j = P_t$ for all j. This leads to the reduced form macro model with three equations.

The first equation relates money demand and with interest and income: an LM-type equation.

$$\frac{w'(M_t/P_t)}{u'(y_t)} = \frac{i_t}{1+i_t} \Rightarrow M_t/P_t = L(y_t, i_t) \qquad \text{LM}$$

The second equation is reminiscent of an IS equation

$$\frac{u'(y_t)}{P_t} = \beta\,(1 + i_t E)\,\frac{u'(y_{t+1})}{P_{t+1}} \qquad \text{IS}$$

The third equation is the aggregate supply relationship

$$v'\,(y_t)\,/u'\,(y_t) = \frac{\theta-1}{\theta} \qquad \text{AS}$$

We see that aggregate supply is constant and independent of monetary policy. This is due to the assumption of price flexibility made earlier. The first two equations of this model are not influenced by that assumption and will still hold later on, when we change it.

The micro-foundations of this model allow us to introduce shocks to different parts of the economy in a straightforward way (rather than tucking them on to the reduced form equations). We introduce three types of shocks:

- ξ_{1t}, shocks to fiscal policy, added to the function u.
- ξ_{2t}, preference or technology shocks, added to the function v.
- ξ_{3t}, shocks to the technology of transactions, added to the function w.

Assuming the shocks and their effects on the endogenous variables are reasonably small, we can rewrite the model in a log-linear approximation

$$\log M_t - \log P_t = \eta_y \ddot{y}_t - \eta_i \ddot{r}_t + \nu_{1t} \qquad \text{LM}$$

$$-\sigma\ddot{y}_t = \ddot{r}_t + E_t\left(-\sigma\ddot{y}_{t+1} - \pi_{t+1}\right) + \nu_{2t} \qquad \text{IS}$$

$$\ddot{y}_t = y_t^S \qquad \text{AS}$$

The notation is as follows: $\ddot{y}_t$ is $\log(\text{Y}_t/\text{Y}_t^*)$, and $\ddot{r}_t$ is $\log(1+i_t)-\log(1+i^*)$, where the values with a star are the equilibrium solutions. Further, π_t is the inflation rate $\log(P_{t+1}/P_t)$ and σ the curvature of the utility function u at $\mathrm{Y}_t = \mathrm{Y}_t^*$.[3] The disturbances are assumed to show up as the terms ν_{1t} and ν_{2t}, and the AS equation shows that output is purely a function of the disturbances.

2.2 Frictions in price setting

We now follow Calvo (1983) by assuming the following kind of friction in the setting of prices: rather than letting all producers change their prices every period, a price may be changed with probability $1-\alpha$. With probability α, the price must stay the same as last period.

When a producer is given the opportunity to change the price, the price that is set is not necessarily the same as the optimal price in the frictionless model. This is because there is a positive probability α that next time, the price cannot be changed, so that it holds for more than one period. Thus the producer will want to maximize an expression like

$$\max_{P_t^j} E_t \left\{ \sum_{k=0}^{\infty} \alpha^k \beta^k \Pi(P_t^j, P_{t+k}, \mathrm{Y}_{t+k}, \lambda_{t+k}, \xi_{2,t+k}) \right\}$$

where Π is the 'profit'-function.[4] Log-linearizing the above expression around $\mathrm{Y}_t = \mathrm{Y}_t^*$ and ploughing through a fair amount of algebra, we obtain a new AS equation, which looks like

$$\begin{aligned} \pi_t &= \kappa \left(\ddot{y}_t - \ddot{y}_t^S \right) + \beta E_t \pi_{t+1}, \text{ where} \\ \kappa &= \frac{(1-\alpha)(1-\alpha\beta)}{\alpha} \left(\frac{\omega+\sigma}{1+\omega\sigma} \right) > 0 \end{aligned}$$

and where ω is the curvature of v at $\mathrm{Y}_t = \mathrm{Y}_t^*$ and $\ddot{y}_t^S$ is the solution for $\ddot{y}$ for the flexible-price case. We can now write a third version of our macro model, subject to price rigidities.

$\log M_t - \log P_t = \eta_y y_t - \eta_i r_t + \nu_{1t}$	LM
$\ddot{y}_t = E_t \ddot{y}_{t+1} - \frac{1}{\sigma}[\ddot{r}_t - E_t \pi_{t+1}] + g_t$	IS
$\pi_t = \kappa \left(\ddot{y}_t - \ddot{y}_t^S \right) + \beta E_t \pi_{t+1}$	AS

We have slightly rewritten equation IS, and named the modified disturbance term g_t. Taking the real side of the economy, IS and AS, we can now show the use of monetary policy rules.

[3] This curvature is given by $-u''(\mathrm{Y}_t^*)\mathrm{Y}_t^*/u'(\mathrm{Y}_t^*)$.

[4] Of course these yeoman-farmers do not think of profit in the proper sense, but they do maximize something similar.

2.3 A policy rule in the model

The system AS-IS above can be amended with a policy rule for the interest rate,

$$\ddot{r}_t = \phi_\pi \pi_t + \phi_y \ddot{y}_t.$$

Using this rule, we can get rid of the $\ddot{r}_t$ in IS. Defining $\mathbf{z}_t = [\pi_t, y_t]'$, $\mathbf{s}_t = [g_t, y_t^s]'$ and substituting in the rule, we have a system that may be represented as

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_t + \mathbf{B}E\mathbf{z}_{t+1} + \mathbf{C}\mathbf{s}_t. \tag{9}$$

By invoking rational expectations, we can write

$$\begin{aligned} \mathbf{z}_t &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}E\mathbf{z}_{t+1} + (\mathbf{I} - \mathbf{A})^{-1}\mathbf{C}\mathbf{s}_t \\ &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{C}\sum_{i=0}^{\infty}\left((\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\right)^i E_t\mathbf{s}_{t+i} \end{aligned}$$

where the last step holds if $\left\|(\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\right\| < 1$. Assuming a VAR process for the shocks, this gives a unique bounded solution for the system. This solution shows how the economy will dynamically respond to a series of shocks $\{\mathbf{s}_t\}$.

An important point needs to be made here. Every time the monetary policy rule is substituted in we must have assumed that the rule is known and credible; otherwise it does not work. Would the central bank take the same actions by surprise, the effects would be different. Thus, precommitment by the bank is crucial for the outcome.

3 A ranking criterion

With the tools we have developed, we can analyze the workings of the economy under different policy rules. To compare the desirability of the different outcomes, a welfare criterion must be specified. Because of the micro-foundations of our model, this criterion can be found relatively straightforwardly. We could, for instance, look at the maximand of the consumers in formula (2).

Because we will be looking at steady states, an equivalent measure is found by looking at the expression for one period,

$$E\left[u\left(c_t;\xi_{1,t}\right) - \int_0^1 v\left(y_t^z;\xi_{2,t}\right)\mathrm{d}z + w\left(\frac{m_t}{p_t};\xi_{3,t}\right)\right].$$

Given that consumers are identical all arguments are aggregate variables. In equilibrium, $c_t = y_t^z$ for all z.

A number of assumptions go into the construction of a criterion from this expression. First of all, the m_t/p_t-term is neglected, focusing attention on the real side of the economy. Furthermore, we will look at only a second order Taylor expansion around a steady state with zero disturbances. A number of other assumptions will follow during the construction.

We start with the terms involving $u(\cdot)$. Define $\hat{y} = \log(y_t/y^*)$, then it is possible to write the expansion as

$$\begin{aligned} u\left(y_t;\xi_{1,t}\right) &= u\left(y^*,0\right) + u_c y^* \left(\hat{y}_t + \frac{1}{2}\hat{y}_t^2\right) + u_\xi \xi_{1,t} + \frac{1}{2}u_{cc}y^{*2}\hat{y}_t^2 \\ &\quad + u_{c\xi}y^*\hat{y}_t\xi_{1,t} + \frac{1}{2}u_{\xi\xi}\xi_{1,t}^2 + O\left(\left\|\xi_{1,t}^3\right\|\right) \end{aligned}$$

From this, we can ignore the terms that are independent of policy (t.i.p., terms that do not change as the policy rule is changed), and take expectations to find

$$\begin{aligned} E(u) &= u_c y^* \left[E(\hat{y}) + \frac{1}{2}(1-\sigma)E(\hat{y})^2 + \frac{1}{2}(1-\sigma)\operatorname{var}(\hat{y}) + \sigma\operatorname{cov}(\hat{y},\bar{c}_t)\right] \\ &\quad + \text{ t.i.p.} + O\left(\left\|\xi^3\right\|\right) \end{aligned}$$

where $\bar{c}_t \equiv -u_{c\xi}\xi_{1,t}/u_{cc}y^*$.

A similar computation can be made for the second term, involving $v(\cdot)$, which comes out as

$$\begin{aligned} & v_y y^*[E(\hat{y}) + \frac{1}{2}(1+\omega)E(\hat{y})^2 + \frac{1}{2}(1+\omega)\operatorname{var}(\hat{y}) - \omega\operatorname{cov}(\hat{y},\bar{y}) \\ & + \frac{1}{2}\left(\theta^{-1}+\omega\right)E\left\{\operatorname{var}_z\left(\log y_t^z\right)\right\}] + \text{ t.i.p.} + O\left(\left\|\xi^3\right\|\right) \end{aligned}$$

where $\bar{y}_t = -v_{y\xi}\xi_{2,t}/v_{yy}y^*$.

Two further assumptions now facilitate the summation of these two terms. First of all, assume that $u_c = v_y$. This means that the steady-state level of output is always efficient, in spite of the shocks. This may be caused by the use of other instruments. Secondly, assume that other policy instruments are adjusted such that $E(\hat{y}) = 0$ regardless of monetary policy. This greatly simplifies the criterion, which may now be written as

$$\begin{aligned} \mathcal{W} &= -\frac{u_c y^*}{2}\left[(\sigma+\omega)\underbrace{\operatorname{var}(\hat{y}-\hat{y}^s)}_{\text{Variability of the output gap}} + \left(\theta^{-1}+\omega\right)\underbrace{E\left(\operatorname{var}_z\left(\log y_t^z\right)\right)}_{\text{Dispersion of output levels}}\right] \\ &\quad + \text{t.i.p.} + O\left(\left\|\xi^3\right\|\right) \end{aligned}$$

The term that quantifies the impact of the dispersion of output levels is only invoked when there is some sort of price rigidity causing the dispersion. Until now we have not made any assumptions about the way in which prices are set. We will use the Calvo-mechanism we have used before to allow for some rigidity. This means that each period, a fraction α of the producers cannot change their price.

The appearance of $\mathcal{W}$ changes somewhat after the Calvo-rule is used. It now reads

$$\mathcal{W} = -\frac{u_c y^*}{2}\frac{\alpha}{(1-\alpha)^2}\left(\sigma^{-1}+\omega\right)\left[\mathcal{L}+\pi^{*2}\right] + \text{ t.i.p. } + O\left(\left\|\xi^3\right\|\right)$$

Here, $\mathcal{L}$ may be viewed as the actual stabilization loss, and π^{*2} as the average rate of loss. It turns out that

$$\mathcal{L} = \frac{(1-\alpha)^2}{\alpha} \left(\frac{\sigma + \omega}{\theta^{-1} + \omega} \right) \operatorname{var} (\hat{y} - \hat{y}^s) + \operatorname{var} (\pi) .$$

The result that the loss function ultimately depends on a weighted average of the variation in output and the variation in inflation is not unfamiliar in macroeconomics. It has often been assumed without a micro-foundation that such a criterion could be relevant. The advantage of this form is that the weights of the two variations are given by the model, rather than picked by the economist.

Note that we started out by omitting losses due to monetary fluctuations, looking only at production and consumption. However, the criterion that we found shows that there still is an aversion to inflation (both in level and in variation).

4 Feasibility of policy rules

Now that we have a criterion by which we can rank the different rules, we might want to look for the rule that minimizes it. This would be the rule that caused zero inflation in every period. However, we must take into account that only a limited amount of information about the economy is known, so that in practice it may not be possible to formulate a rule that produces $\pi = 0$.

We again post a version of our macro model, leaving out the LM equation[5] but adding a policy rule.

$$\pi_t = \kappa \tilde{y}_t + \beta E_t \pi_{t+1} \qquad \text{AS}$$

$$\tilde{y}_t = -\tfrac{1}{\sigma} [r_t - E_t \pi_{t+1} - r_t^n] + E_t \tilde{y}_{t+1} \qquad \text{IS}$$

$$r_t = \phi_\pi \pi_t + \phi_y \tilde{y}_t + q_t \qquad \text{PR}$$

where $\tilde{y}_t = y_t - y_t^s$ is deviation in output and $r_t^n = \sigma \left(g_t + E_t y_{t+1}^s - y_t^s \right)$ is the combined effect of all disturbances. The rule PR allows for an exogenous shock q_t by the policy makers.

[5] The LM equation determines the demand of money but does not influence the real economy in this model.

We can now explicitly write down a system such as (9). There holds that

$$\begin{aligned}
\begin{pmatrix} \pi_t \\ \tilde{y}_t \end{pmatrix} &= \tilde{\mathbf{B}} \begin{pmatrix} E_t\pi_{t+1} \\ E_t\tilde{y}_{t+1} \end{pmatrix} + \tilde{\mathbf{C}} \begin{pmatrix} q_t \\ r_t^n \end{pmatrix} \\
&= \frac{1}{\Delta} \begin{pmatrix} \frac{\beta\sigma+\kappa+\beta\phi_y}{\sigma} & \kappa \\ \frac{1-\beta\phi\pi}{\sigma} & 1 \end{pmatrix} \begin{pmatrix} E_t\pi_{t+1} \\ E_t\tilde{y}_{t+1} \end{pmatrix} - \frac{1}{\sigma} \begin{pmatrix} \kappa \\ 1 \end{pmatrix} (q_t - r_t^n) \\
\Delta &= \frac{\sigma + \kappa\phi\pi + \phi_y}{\sigma}
\end{aligned}$$

From here we can answer a number of important questions. One question might be whether the economy has a determinate rational expectations solution. The answer lies in the properties of matrix $\tilde{\mathbf{B}}$: if both eigenvalues lie inside the unit circle such an equilibrium exists. It can be shown that this condition is satisfied if $\phi_\pi > 1$ and $\phi_y \geq 0$; this is not a necessary condition, however.

Another question, given that a solution exists, might be what the effect of the original shocks ξ_1, ξ_2 and ξ_3 would be on output and inflation. The answer to that question can be found by examining the effect of the ξ's on r_t^n and looking at $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$.

The effects of exogenous shock q_t can also be derived from this system. It appears that a policy strategy that sets $q_t = r_t^n$ for all t would result in the minimum possible loss criterion, for all values of ϕ_π and ϕ_y that generate a determinate solution. The practical feasibility of such a policy is questionable, however.

5 Quantification and counterfactuals

Now that the model is completed, professor Woodford showed how we can analyze real-world monetary policy in terms of policy rules, and try how alternative rules might have fared under the shocks that actually occured. The methodology runs as follows:

1. Estimate an unrestricted VAR model of the interest rate, the inflation rate and output.

2. Choose the parameters of the structural model so that the predicted response to monetary policy agrees with the VAR estimate.

3. Using the quantitative specification of the structural model, and the VAR to model the expectations, identify the shocks that actually occured by the residuals of the equations when fit to the data.

4. The parametrized model and the shocks can now be used to simulate the consequences of counterfactual monetary policy rules for the evolution of $\{r_t, \pi_t, y_t\}$.

5. Using the criterion derived in section 3, evaluate the welfare consequences of alternative policy regimes.

The period over which the analysis is conducted is 1979–95. The model that is used is a little different from the one described above, incorporating a two-period decision lag for goods purchases. This adaption was made to accomodate the data. Also, there are decision lags in pricing by suppliers, which differ across goods. Without these adaptions, shocks would affect the real economy almost immediately; this is contradicted by the data.

We already noticed that a rule that causes zero inflation would be infeasible, because there is not enough information to 'fuel' such a rule. We could solve for the best value of $\mathcal{W}$ given that the rule can only depend on known variables. However, it turns out that such a rule will cause negative interest rates at certain points in time: a practical impossibility.

To avoid this problem, we can impose a condition that adds a penalty for interest rate variability to the loss-criterion. The optimal rule that comes out this problem is indeed feasible, but not very practical as it is very complicated (imagine how it uses *all* available information to come up with the best interest rate).

The rules that validate real interest are those that are simple in structure, yet imply a value for the loss criterion that is close to optimal. In Rotemberg and Woodford (1998), three such rules are discussed:

1. A generalized Taylor rule: $r_t = a\pi_t + by_t + cr_{t-1}$.
2. Dependence on lagged data: $r_t = a\pi_{t-1} + by_{t-1} + cr_{t-1}$. This is the above rule in the case where data becomes available with a lag.
3. Price level targeting: $r_t = aP_t + by_t + cr_{t-1}$. This rule can also be applied in a 'lagged' version.

In each case the question is what the optimal parameters a, b, and c are, whether the rule generates determinate paths for the state variables, and how much the loss criterion is minimized. We look only at the rules in case *1*, the others may be found in the paper.

For rules in category *1* with c set to zero, the optimal values are $a = 2.88$ and $b = 0.02$. This is different, but not shockingly so, from Taylor's (1993) values. Rotemberg and Woodford (1998) show how the loss criterion behaves for different a's and b's in a diagram. The value of the loss criterion gets very close to that attained in the optimal case discussed above. Rules in category *1* with c unequal to zero include a feedback mechanism in r_t. Surprisingly, diagrams show that the rule generates stable paths for values of c as high as 10! This is due to the fact that the rule is a *promise* to raise interest rates, which is not necessarily carried out. The optimal value is close to the previous one: $a = 1.22$, $b = 0.06$ and $c = 1.28$.

For the above optimal rules, the counterfactual loss criterion is lower than it was in the real case. This suggests that something can be gained studying these optimal rules.

6 Conclusions

At the end of the lectures, a number of important conclusions could be drawn. The most striking result probably is that a simple (backward-looking) Taylor-type rule can achieve outcomes that are nearly as good as those achievable by any policy. This suggests that those rules can be a valuable yet inexpensive tool for central banks.

It is often suggested that a central bank's policy should be forward-looking, that the bank should carry out 'preemptive strikes' before a problem actually occurs. This model shows that such is not necessary, as long as the economic agents themselves are forward looking. The promise of the central bank to raise interest rates in certain cases, laid down in a Taylor rule, is enough to stop those cases from happening. The credibility of such a promise is, of course, essential.

Because data often takes time to reach the economist, it is reassuring that the performance of the rules is but weakly affected by the use of lagged data.

Rules that specify short-term interest rates should incorporate the lagged endogenous variable as a regressor, with a coefficient larger than one. Because of rational expectations, this will not result in instability of the economy. There is little gain from making the rule dependent on realized values of aggregate production.

Of course all these conclusions stand or fall with the faith that one puts in the simple macro model that they come from. However, the ability of the model to mimic real data is encouraging, as is the fact that Taylor rules can very well explain past monetary policy. Thus, applying them in the future should not be a step in the dark.

The lectures that professor Woodford gave in Wageningen were by no means 'an easy ride,' but with this impressive list of conclusions well worth the effort. Future data will show whether this work has had impact on the workings of, say, the European central bank.

7 References

- Calvo, G., 1983, "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, **12**(3):383–98.

- Rotemberg, J.J. and M. Woodford, 1992, "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity," *Journal of Political Economy*, **100**(6):1153–1207.

- Rotemberg, J.J. and M. Woodford, 1997, "An Optimization-Based Econometric Framework for the Analysis of Monetary Policy," *NBER Macroeconomics Annual 1997*, pp. 297–346.

- Rotemberg, J.J. and M. Woodford, 1998, "Interest Rate Rules in an Estimated Sticky-Price Model," *mimeo*, April.

- Taylor, J.B., 1993, "Discretion vs. Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, **39**:195–214.

QUESTIONNAIRE
UTRECHT-COURSES 1999 - 2000

Please circle the number(s) of the course(s) you wish to follow next year and return the form before **16 May 1999** to the NAKE Secretariat. In principle a course will only be scheduled once every two years. The courses marked with a star (*) have been given in the academic year 1998-99 and will therefore not **normally** be available in the academic year 1999-2000. Courses in bold face are new (i.e. these have been proposed recently but have not yet been scheduled).

10-week courses (4 SP = 160 hours)

	Teacher(s)	*Institute*	*Course*
1	van den Berg/van Ours/ den Butter	VU/KUB	Applied labour economics
2*	Peters/Borm	MU/KUB	Game theory
3	Hartog/Teulings/Theeuwes	UVA/EUR	Advanced labor economics
4*	Palm/Nijman	MU/KUB	Theory and application of modeling volatility in financial economics
5*	Wansbeek	RUG	Econometrics of panel data
6	Talman/van der Laan	KUB/VU	General equilibrium model
7	Folmer/de Zeeuw/	LUW/KUB	Environmental problems and policy: A) A theoretical introduction
	Withagen/Smulders	VU/KUB	B) Growth and environment

5-week courses (2 SP = 80 hours)

8	......	..	....
9	Beetsma	MinvEZ	Topics in international macroeconomics
10	van Bergeijk c.s.	DNB	Applied policy analysis
11*	Bovenberg	KUB	Fiscal policy in open economies
12	Brakman/Van Marrewijk	RUG/EUR	Regional economics, agglomeration and the global economy
13	Brenner	RUU	A critical view of economic theory
14	Brenner	RUU	Social economics: Heterodox approaches to economic theory
15*	Broer	CPB/EUR	Dynamic general equilibrium modelling
16	Burrell/Oskam	LUW	Agricultural policy analysis
17	den Butter	VU/WRR	Macroeconomic policy modelling
18	Cukierman	KUB	Central Bank strategy, credibility, and independence
19*	van Damme	KUB	Topics in applied microeconomics: Variable topic
20	van Dijk/Boswijk	EUR/UVA	Econometric inference in dynamic models with integrated processes
21	Ellman	UVA	The political economy of transition
22	**Ellman**	**UVA**	**The economics of famines**
23	van Ewijk/van Wijnbergen	CPB/UVA/EZ	Economic growth and development: Macroeconomics
24	van Ewijk/Oosterbeek	CPB/UVA	Economics of education
25	Furth/Van Cayseele	UVA/KL	Advanced industrial organisation
26*	Garretsen/Sterken/Van Ees	KUN/RUG	Capital market imperfections, investment and monetary policy
27	de Gooijer/Franses	UVA/EUR	Recent developments in non-linear time series analysis
28	Goyal/Janssen	EUR	Coordination problems
29	de Gijsel	MU	Micro-economische onderbouwing van een monetaire economie
30*	Gunning/Keyzer	VU	Current issues in development economics
31	Hartog/Theeuwes	UVA	Labour economics: A comparative empirical perspective
32	Heijdra/Meijdam	RUG/KUB	Intertemporal aspects of macroeconomics
33	Heijdra	RUG	New Keynesian macroeconomics

34	**Heijdra**	**RUG**	**The macroeconomics of monopolistic competition**
35	Herings	KUB	Theory of incomplete markets
36*	Hommes	UVA	Nonlinear economic dynamics
37	Houba	VU	Differential games in economics
38	**Houba**	**VU**	**Strategic bargaining and endogenous threats**
39	Huizinga	KUB	International factor movements and international financial markets
40*	Jepma	RUG/UVA	International environmental policies
41	de Jong, E.	KUN	Exchange rate economics
42	de Jong, F.	KUB	Economics of foreign exchange markets
43*	Keyzer	VU	Applied general equilibrium models
44*	Kloek	EUR	Visualising data
45	Kooreman/Kapteyn	RUG/KUB	Intertemporal choice
46	Kooreman	RUG	The economics of household behaviour
47	van der Laan/Talman	VU/KUB	Economic equilibrium under price restrictions
48	**Magnus**	**KUB**	**Sensitivity analysis in econometics**
49*	Magnus	KUB	Optimisation methods in econometrics
50*	Maks	MU	Competition and market coordination
51	Meijdam/Verbon	KUB	Theories of government debt
52	Morgan	UVA	History of economic ideas
53*	Morgan	UVA	History and philosophy of economic models
54	**Muysken**	**MU**	**Low skilled unemployment, job competition, and overeducation**
55	**Nijkamp**	**VU**	**Meta-analysis in economics**
56	Peters/Storcken	MU	Social choice theory
57	Pfann	MU	Optimal investment contingency plans of firms
58*	Potters	KUB	New institutional economics
59	**Potters**	**KUB**	**Market micro-structure**
60	**Reuten**	**UVA**	**Heterodox economics: Marxian political economy**
61	**Ruys**	**KUB**	**Regulation and privatisation**
62	**Ruys**	**KUB**	**Competition and cooperation in the non-profit sector**
63*	Schoonbeek	RUG	Topics in oligopoly theory
64	Schoorl	RUG	History of Dutch economic thought
65	Schram/van Winden	UVA	Experimental economics and the design of mechanism
66	van Soest/Melenberg	KUB	Applied non-parametric and semi-parametric econometrics
67	Steenge	UT	Rational choice theory and governance in the public sector
68*	Thijssen/Goudriaan	LUW/IOO	Efficiency and productivity analysis
69	Uhlig	KUB	Business cycles
70	Verbon	KUB	Decision-making on intergenerational transfers
71	Vorst	EUR	Options pricing theory
72	Vorst/van de Sar	EUR	Behavioral Finance
73	de Vos	VU	Bayesian views on testing and model selection
74*	de Vries/Lucas	EUR	Financial risk management
75	**de Vries**	**EUR**	**Advanced monetary economics**
76*	Wakker	LU/KUB	Prospect theory
77	Wansbeek	RUG	Latent variables and methods of moments estimation
78	**Wansbeek**	**RUG**	**Econometric methods in marketing**
79	Weddepohl	UVA	Overlapping generations models
80*	van Winden	UVA	Behavioural modelling of government decision-making
81	van Wijnbergen	UVA/MinvEZ	Economics of transition
82*	Smulders	KUB	Endogenous growth theory
E01	**Sterks**	**RUG**	**Law and economics**
E02	**Ziesemer**	**MU**	**Topics in international trade**
E03	**Balder**	**UU**	**Economies with a continuum of agents**
E04	**Rietveld**	**VU**	**Transport economics**
E05	**Oosterhaven**	**RUG**	**Infrastructure and economic development**