Forecasting Inflation With a Nonlinear Output-Gap Model

Robert J. T. Hillman

24th November 1998

Abstract

I examine the extent to which modelling a simple form of nonlinearity in the output-gap/inflation relationship is useful for one-step-ahead out-of-sample forecasting of the quartlerly US inflation rate. If nonlinearity is an important feature of this relationship there are strong policy implications. Previous work has focussed on in-sample testing, but as is well known, in-sample evidence for nonlinearity is not as convincing as evidence based on out-of-sample inference. As the output-gap is unobservable I compare three different estimates of the output-gap for their forecasting ability. Two of the measures turn out to have significant value from a forecasting perspective, and furthermore there is convincing evidence that the output-gap should enter nonlinearly in the forecasting model.

1 Introduction

There is little consensus as to what models are best for forecasting inflation. A number of economic theories do exist that suggest exploitable relationships between fundamental variables and the inflation rate. Some examples include mark-up pricing, cost-push and simple Phillips curve type relationships. However, whilst certain economic models have been seen to be useful at times, no single general model seems to have been discovered (Stockton and Glassman's study is seminal on this point[26]) that is consistently preferable. There have been two main ways that economists have approached this current state of the literature with a view to improving the situation.

One way is to recognise that in practice one rarely relies on one model forecast, but rather some sort of consensus forecast is arrived at, often using a mixture of subjective adjustments and statistical forecast combination. The latter is motivated by the commonly observed stylized fact that simple combinations (simple or weighted averages) of forecasts from different sources can give better forecasts than any individual forecast. One recently proposed development in forecast combination is to allow for different forecasts to have different weights in a combination over time, see for example [9] and [16]. Such methods do seem able to produce improved forecasts, and in the latter work the methodology offers insights into why different models may be more appropriate at different times[1]. The essence of the combination approach is that there may be different statistical relationships through time, and individual estimation of different models, followed by a pooling of the models is one way to capture this.

A second approach to the problem of forecasting inflation has been to generalise the form of statistical relationships that the earlier models allowed. All of the models examined by Stockton and Glassman were linear. However there are frequently claims made that economic variables display nonlinearity, particularly in their univariate time series behaviour. Clearly a natural response to such detections of nonlinearity is to examine whether allowing for potential nonlinearities between say the money supply, output-gaps, lagged inflation and so on, can actually improve forecasting. Unfortunately economic theory does not often offer much guidance to the precise form of nonlinearity that may be useful to model, and nonlinear model specification methods are little developed[2]. As a result most of the nonlinear forecasting attempts have used a common strategy. This is to impose virtually no restrictions on the form of nonlinearity at all, and to search for the best fitting model, typified by the application of neural networks for example. The problems of this approach such as overfitting, approximation of spurious relationships, lack of parameter identification, and the ensuing poor inferential ability are well known. As I will review in the next section, the forecasting results of these applications has been disappointing.

In this paper I investigate the forecasting ability of some very specific forms of nonlinearity. Whilst therefore I may not be able to model precisely some genuine nonlinear relationship in the data by restricting the functional form as much as I do, I greatly reduce the optimization, inference, and overfitting problems associated with more general nonlinear models. What is more the precise form of nonlinearity that I look for is supported and suggested by economic theory, and some recent associated empirical work. There are also strong policy implications suggested by the use of these restricted functional forms, the estimated relationships are interpretable from a policy perspective. The particular form of nonlinearity I study is also closely related to the recent work on forecast combination. The essence of the approach is the same, that is, there may be different statistical relationships between sets of variables over

[1] An interesting methodological implication suggested by following this line of research is that the goal of finding a single global (i.e. a model that works all the time) is misguided. Rather we should allow for a diversity of models, recognising each models strengths (and weaknesses, for example over time or in different circumstances), and concentrate on developing methods to discrimininate between models over time. Another example of work in this vein is Chen, McCulloch and Tsay (1996 [3]).

[2] I mean nonlinear in the widest possible sense, which implies a vast set of functional forms. There are a number of specification methods that have been developed for use within specific classes of nonlinear model, such as threshold models and Markov-switching models. These methods are generally 'specific-to-general' within a class.

time.

A result is that I do find convincing evidence (far stronger than the results provided by researchers using more flexible models) that modelling nonlinearity is useful for forecasting inflation, at least within the set of variables I examine. At the end of paper I discuss some more philosophical issues as to what the estimated relationships in this paper can allow us to say about the nature of the 'true' model generating the data we observe.

The rest of this paper is organised as follows. Section 2 contains a brief review of some recent attempts to examine the usefulness of modelling nonlinearity in forecasting the inflation rate. Section 3 begins with a review of some micro-foundations that suggest the output-gap/inflation relationship will be nonlinear, and then proceeds to summarise the testing methods and extant evidence. In Section 4 I introduce my modelling procedure, that combines adaptive modelling, automatic variable selection techniques and threshold modelling. Section 5 describes the data, and an important issue, that of how to estimate the output-gap measure itself. Section 6 reports details on the model estimates, and Section 7 contains the forecasting results. In Section 8 I give a brief discussion of how to interpret the results, and in particular what, if anything, we can infer about some true DGP based on our results. Section 9 concludes with a discussion of some limitations of this paper, and some suggestions for future work.

2 Nonlinear Forecasting Models of Inflation

As mentioned above, there has been some work on nonlinear forecasts of inflation, but unfortunately the results of these exercises have not been very enlightening as to the form of nonlinearity that might be important. For example Moshiri, Cameron and Scuse ([21] 1997) and Cameron and Moshiri ([2] 1997) have explored neural network based models of inflation. A generic neural network model of inflation would have the form

$$\pi_t = \theta' Z_t + \sum_{i=1}^{j} \lambda_i g(Z_t; \psi_i) + e_t$$

where θ, λ_i ($i = 1$ to j), and ψ_i ($i = 1$ to j) are parameters. Z_t is a vector of inputs, and $g()$ is a nonlinear function. If Z_t contains lagged inflation and an output gap term then clearly we are in effect estimating equation (1) as the first part of the above, but now augmented by a number of nonlinear terms. Strictly speaking the model given above is a semi non-parametric model rather than what we would normally think of as a neural network, the latter would not contain the separate linear part. The choice of $g()$ can be arbitrary[3]. We know though that for some simple squashing functions such as sigmoidal functions, given sufficiently enough of

[3] This form is also that used by Lee, White and Granger (1993 [19]) as a means of testing for nonlinearity in mean.

them ($g()$ functions on the right hand side i.e. j is large enough) this model can approximate a wide range of nonlinear functions ([28] 1989). Moshiri, Cameron and Scuse ([21] 1997) and Cameron and Moshiri ([2] 1997) estimated models like the one above, **without** the linear part, using a variety of squashing functions (and also a slightly more complicated version that contains feedback from the $g()$ function outputs to the Z_t vector). The dependent variable was Canadian CPI inflation. They used as inputs combinations of lagged and expected inflation, money supply and an output gap term. Overall they found some weak evidence for nonlinearity. Though no formal tests for statistical significance were carried out they concluded that some of the neural models they estimated outperformed the linear benchmark comparison models at some horizons in terms of standard criterion like mean square error and mean absolute error. The results were however sensitive to the form of $g()$ functions, some outperformed the linear model at a particular horizon whilst other nonlinear models were not as accurate.

Chen and Swanson (1997 [4]) looked at one-step-ahead forecasts of US inflation. They looked at a range of neural network based model with different $g()$ functions with and without the linear part. They found that in terms of mse a purely nonparametric model (i.e. without the linear part) outperformed all other models including linear models and a random walk specification. The chosen model used lagged inflation and lagged output gap terms as regressors. However the success of this model was not overwhelming, it did not significantly outperform a linear model with the same set of explanatory variables on the right hand side. Also the linear specification with lagged inflation and a single lagged output gap term was best in terms of mean absolute error (mae) and in terms of predicting direction. A further problem in attaching any significance to these weak forecast improvements is that the nonlinear model that ex post was found to be best by mse was one of 28 nonlinear models fitted. Many of the other 27 models performed very badly (many significantly worse) compared with the linear model. Finally the form of the nonlinearity captured by this model is not obvious. The 'black-box' nature of the neural network model makes learning anything about potential structural nonlinear relationships very difficult.

In many ways the results of this body of work are not particularly surprising. It is well known that optimization of neural network type models is inherently difficult, mainly due to the lack of identification within the model and the dual problem of both estimating the parameters and the network size (the number of hidden units). Furthermore such an approach is intrinsically limited to being able to provide at best improved forecasts, these models would not be able to tell us much about the form of nonlinearity, it's interpretation and economic significance.

3 What Theory and Extant Empirical Work Has to Suggest About Nonlinearity in the Determination of Inflation

3.1 Microfoundations

The ex ante case for nonlinearity between the output-gap and inflation is quite strong. Unfortunately, it is not unambiguous as to the precise form of nonlinearity. Two recent papers deal very comprehensively with the micro-foundations that might lead one to believe the relationship is nonlinear, Yates ([29] 1998) and Dupasquier et al ([11]1998). I will give three examples that give the flavour of these arguments. Firstly, and probably the most obvious argument is based on capacity constraints. One would expect that with increasing marginal costs and fixed capacity in the short run, some firms find it costly to increase their capacity in the short run. This would mean inflation becomes more sensitive to output in times of excess demand, implying a convex shape, consistent with the original Phillips curve idea. A second form of nonlinearity would be implied by a signal extraction problem as proposed by Lucas ([20] 1972). In this world agents have to try and untangle relative price movements from aggregate price movements. Agents would only want to respond to changes in relative prices ideally. However the more volatile is inflation the less agents will be able to filter out relative changes, and so the less will be their output response. Under conditions of high inflation then the slope of the relationship will be steeper than under times of low inflation. The nonlinearity here is a form of state-dependence, where the slope is a function of the volatility of inflation. A third and final argument that implies nonlinearity is based on mononpolistic competition. If the economy is made up of such firms, then producers may be inclined to lower prices quickly to avoid being undercut. However they may be reluctant to raise prices in the hope of keeping out new competitors. This would imply a concave Phillips curve, see Stiglitz for an argument for this case ([25] 1984). Faced with these different and conflicting theoretical arguments a recent literature has developed that attempts to discriminate between these models empirically.

3.2 Empirical Evidence

As discussed in the previous section, there have been recent arguments proposing the relationship is linear [13], concave [25] and convex [5]. There are strong and different policy implications depending on which view is right. The previous authors all discuss the policy issue thoroughly and I will not repeat all the arguments here, but one of the most obvious examples of why this is important is concerned with the timing of policy. If the relationship is convex that means that if the economy shows signs of overheating it may well be beneficial to act as quickly as policy to dampen the economy. The reason is that in order to maintain a stable inflation rate the economy has to spend proportionately longer in a disinflationary regime (with unemployment higher than the natural rate) that it does in an expansionary or overheating regime. The reason why the empirical evidence for the form of nonlinearity is so important is that the opposite is true if the relationship is concave! Next I will describe the methods and results that exist in the literature for testing for the form of this relationship.

The simple linear model posits that inflation is driven by inflation expectations and the output gap (or the unemployment gap)

$$\pi_t = \pi^e_{t+1} + \beta \ ygap_t + e_t \tag{1}$$

where π is inflation, π^e inflation expectations and $ygap$ the output gap term. e_t is a stochastic error. There are a number of ways the output gap may be measured, and later I will go into more detail. For now consider it measures the difference between actual output and potential output. The inflation expectations terms presence in this simple equation suggests that this is the main factor thought to drive inflation. This can be modelled in a number of ways, popularly, by dividing the term into a forward and backwards component, so tying inflation down in part to the past. We could write this easily as

$$\pi_t = A(L)\pi_{t-1} + B(L)\pi^e_{t+1} + \beta \ ygap_t + e_t$$

where now the $A(L)$ and $B(L)$'s are lag operators. Lags of inflation might also be expected to enter if there is some intrinsic dynamics induced by overlapping wage contracts or costly price adjustments. Although theoretically unappealing it can be difficult to reject the presence of a unit-root in inflation (i.e. $A(L) + B(L) = 1$) [4].

Testing for nonlinearity has generally proceeded by introducing the possibility of a simple form of asymmetry. The idea is that the effect of the output gap on inflation is dependent on the sign of the output gap. The simplest form of asymmetry suggests that when the output gap is positive (i.e. when output is above the potential rate, perhaps described as boom conditions) the output gap has a greater effect on inflation than when in recessionary times. An obvious way to test for this is to estimate the augmented model

$$\pi_t = A(L)\pi_{t-1} + B(L)\pi^e_{t+1} + \beta \ ygap_t + \gamma ygap^{pos}_t + e_t$$

and test that $\gamma > 0$. If $ygap$ is defined as $y_t - y^p$ (actual output - potential output) and y^p is the level of output attainable on average in a stochastic economy, then we might expect $ygap$ to have a zero mean. This would be true in a symmetric trade-off. With an asymmetric trade-off however when positive gaps have more effect than negative gaps, then for inflation to be bounded if $\gamma > 0$ we would require the $ygap$ term to have a negative mean[4]. In practice then in testing this equation it is important to redefine $ygap \equiv y - y^p + \alpha$, where α is estimated. This type of asymmetry suggests

[4]This argument can be made in the context of a Phillips curve. If the Phillips curve is linear the NAIRU is where the curve cuts the x-axis. This would also be true if the economy was hit

a kinked or piecewise Phillips curve, at output gap levels greater than α the output gap has an increased effect on inflation. This equation has been used by [5] [18] who found evidence in favour of this form of asymmetry.

A more flexible form of nonlinearity that allows for marginal convexity has also been used to test for an alternative to linearity. In this case we might test an equation of the form

$$\pi_t = A(L)\pi_{t-1} + B(L)\pi^e_{t+1} + \beta\ (y^p_t - y_t)/y_t + e_t$$

in which case the coefficient β measures the degree of convexity. Equations of this form (though using the unemployment gap rather than the output gap) have been tested by Clark and Laxton (1997 [6]) and Debelle and Laxton (1997 [8]). The testing procedure has been to compare the fit of this equation with the simple linear version in (1) above. Though there are problems with unrestricted testing of these models (see in particular [8][5]), the evidence seems to be fairly clearly in favour of the nonlinear specification, but varies for different countries. In particular the strongest evidence for nonlinearity is found in the US data, next Canada, then the weakest evidence in the UK data.

4 A Threshold Model Approach to the Forecasting Problem

An alternative method of modelling the potential asymmetry in the Phillips curve relation is via a threshold model. Threshold models have been used for other types of asymmetry in macro-economics with some degree of success already. For example Hansen (1996 [15]) modelled US unemployment and the asymmetric dynamics between recessions and booms. Peel and Speight (1998 [22]) carried out a similar exercise with more unemployment rates. I will consider a simple form of threshold model

$$\pi_t = A(L)\pi_{t-1} + I(ygap_{t-d} \leq y^*)\beta\ ygap_{t-d} + I(ygap_{t-d} > y^*)\gamma\ ygap_{t-d} + e_t$$

where $I()$ is an indicator function.

This model has a common part consisting of lags of inflation. In principle I could model this expectations part more carefully. One potential problem with solely using backward looking expectations is that we would not expect the parameters estimated

by a symmetric distributed shocks. The average unemployment rate would also be the NAIRU. However in a stochastic economy the average level of unemployment when the Phillips curve is convex is higher than the determinstic (and linear) NAIRU. See Clark and Laxton (1997) for a good exposition [6].

[5] The problem is essentially that if we compare a linear and nonlinear model, statistically the linear model will often fit the data better. However the implied variability of the NAIRU from the linear model is too high. This suggests imposing priors on the models, and once this has been done the linear model will appear less attractive in comparison to the nonlinear model.

in $A(L)$ to be invariant to changes in policy over time. For this reason we might want to include some actual forward looking variables like survey measures and so on. In partial mitigation of this problem I can argue that we are going to estimate these models adaptively. I don't therefore in practice require that either the same lags enter $A(L)$ each period, or that the coefficients on the lags are constant. As it turns out the parameter estimates on these variables are very stable.

The coefficient on the *ygap* variable is allowed to be different depending on whether $ygap_{t-d}$ is above or below the threshold level of *ygap*, y^*. The simple Phillips curve suggests $ygap_{t-1}$ should enter the equation (and all the examples in the testing literature that uses kinked functions solely look at this specification) but in practice I view this as an empirical question. I follow a simple procedure to choose d, the lag of *ygap*. First I estimate a range of linear models allowing up to four lags of inflation and one lag of *ygap* to enter the equation. I choose the preferred model as the one which minimizes the SIC[6], given as

$$SIC = \log(mse) + (p \log n)/n$$

In other words I estimate models with no lagged output gap, and then one set of models for each lag of the output gap term (1 to 4) with all possible combinations of 4 lags of inflation The lag of the output gap that minimizes this equation then serves as the d for the threshold extension. One can view this procedure as a sensible means of augmenting the linear model. First I find the best linear model, then I allow an asymmetry to enter that linear model in a very specific way. As in the test for asymmetry explained earlier this threshold model allows for a piece-wise linear approximation to a potentially smooth relationship[7]. The estimate of the threshold parameter y^* controls where on the *ygap* axis the kink occurs. The earlier arguments about the inclusion of the α parameter (which is not needed with this approach) suggest it ought to be where *ygap* is negative, though this is of course an empirical question.

I follow an adaptive modelling approach to forecasting the inflation rate, which means I allow both the model specification (in terms of the selected explanatory variables) and the coefficients to change before making each one-step ahead forecast. The procedure can be summarised as below

1. Start with sample 1959:3 to 1985:3 (see below)

[6] In principle we could use other selection methods such as the Akaike Information Criterion (AIC), Predictive Stochastic Criterion (PSC), MSE, sign-predictability metrics and so on. Pesaran and Timmerman (1996) discuss different methods, and other papers by Swanson and others explore this issue, eg. Swanson and White (1995 [27]). It is perhaps worth noting that unlike the AIC, the SIC is a consistent selection method. This means that if there is a time-invariant DGM, then recursive application of the SIC (assuming the actual regressors are included in the potential set) will lead to the correct model being chosen eventually.

[7] Of course, we have few priors to suggest the Phillips curve is in fact a convex curve. It might be a threshold or piece-wise function. I suspect that we could not satisfactorily discriminate between a convex curve and a 2 or 3 regime piecewise approximation anyway, with the sample sizes available.

2. Fit range of models of the form

$$\pi_t = A(L)\pi_{t-1} + ygap_{t-d} + e_t$$

   where the variables included on the right hand side are reach possible subset of the set of potential regressors $S = (\pi_{t-1}, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}, ygap_{t-1}, ygap_{t-2}, ygap_{t-3}, ygap_{t-4})$. Choose the model that minimizes the SIC, thereby determining the 'optimal' lag selections for inflation and d for the $ygap$ term.

3. Estimate the threshold model, for d and the inflation lag vector that has been selected above

$$\pi_t = A(L)\pi_{t-1} + I(ygap_{t-d} \leq y^*)\beta ygap_{t-d} + I(ygap_{t-d} > y^*)\gamma ygap_{t-d} + e_t.$$

4. Produce one step ahead forecasts from both linear and nonlinear models (e.g. for the first time for 85:4).

5. Add the next observation to the estimation (in-sample set) then go back to 2.

This approach allows for coefficient evolution and for the maximum use of past data before making each one-step-ahead forecast[8]. We obtain 52 genuine out-of-sample one step ahead forecasts.

My interest is in the forecasting potential of the nonlinear model, but nevertheless there is an issue as to whether we would actually choose the nonlinear in real time. There's two ways to look at this problem. Firstly we can ask whether the nonlinear model fits the in-sample data sufficiently better than the linear model. Just how 'sufficiently' is subjective. Conventional wisdom suggests that the fit has to better than the linear fit, even after taking into account the fact the nonlinear model has more degrees of freedom than the linear model. This 'parsimony' criterion is designed to avoid the practice of over-fitting. We could use penalized likelihood measures such as SIC or AIC to perform this. For example Peel and Speight (1998) used AIC to select their nonlinear models over their linear competitors. An alternative approach would be to use a bootstrap likelihood ratio (LR) test. We would need to use the bootstrap because the LR distribution is non-standard due to identifiability problems inherent in the mixture structure of the threshold model, see for example [17] and [14]. The second way to deal with this question is to say, actually it doesn't matter too much about how much better the nonlinear model is in-sample, what really matters is the out-of-sample performance. For example, we might then decide that if our nonlinear model has been outperforming the linear model out-of-sample for 5 years, we should continue using the nonlinear model for the next one-step-ahead forecast and so on. This might be regardless of the in-sample performance of the model. This

[8] It would of course be of interest to look at multi-step forecasts. I leave this to future work.

would be an example of a more general approach to model selection in which we select forecasting models as being useful based on how well we think they might forecast in the future. We would surely want to look other measures than in-sample fit alone to achieve this aim, especially when considering nonlinear models.

5 The Data and Output-Gap Construction

Inflation is defined as $400\log[p_t/p_{t-1}]$, where p_t is the implicit price deflator for GDP. y_t is real GDP. Both series for the US are measured quarterly and drawn from FRED, the Federal Reserve Economic Database[9]. We have data from 1959:3 to 1998:2.

Up to now I have described a relationship between inflation and the output-gap without actually defining the output gap very precisely. The essence of the term is that it captures to what extent the economy may be operating above or below some trend level. In practice we need to define some measure of this trend. I have two criterion that my measure must fulfill. Firstly I want to be able to estimate real-time output-gaps, at least to the extent that I don't explicitly use information later than the time period at which my last in-sample observation occurs. This way I can (almost![10]) argue that the forecasts are genuinely out-of-sample. Secondly, because I want to re-estimate the output-gap measure each period I require a fairly simple method of estimation.

There is a huge literature on estimating output-gaps and potential output. Techniques range from mechanical filters (such as moving-averages of actual log GDP) to full blown macro-models that endogenously estimate potential output. Barrell and Sefton ([1]1995) provide a recent comparison of some of the most popular measures. One result from this survey is that the macro model based estimates are not strikingly different from simple mechanical filters, though the model measures may offer some more insight into the underlying nature of potential output. In this paper I have therefore used three measures that meet my criterion.

The first measure is a rolling linear trend. I follow a similar method to that proposed by Chen and Swanson (1997) to generate this output gap term. This allows us some comparability with their forecast results. We construct the variable $ygap_t = 400(y_t - y^p_{t+1})$, where y^p_{t+1} is a measure of potential output. Given this is unobserved and we only want to use information up to $t-1$ to estimate it and ensure the forecasts of time t are consequently truly out-of-sample, we use a very simple method to estimate y^p_{t+1}. We fit a linear trend to y_t using rolling samples of 30 quarters, and for each rolling sample we forecast one quarter ahead to get an estimate of y^p_{t+1} given $\{y_i\}^t_{i=t-30}$. (e.g. we estimate the model $y^p_t = \widehat{c} + \widehat{d}t$, then forecast y^p_{t+1}). Whilst

[9] Available over the internet at http://www.stls.frb.org/fred/data/gdp.html.

[10] Strictly speaking, even doing this does not mean the information I use to make the forecast at say 85:3 was actually available at 85:3. Some data are published with lags, and also subject to revisions at later dates. A number of people are working on the construction of genuine real-time data sets (Recent surveys can be found in Ghysels et al (1997), and Diebold and Rudebusch (1991))..

the linear deterministic trend is in the long-run clearly inappropriate, the rolling estimation does produce intuitively rationalisable output-gap series. I attach the prefix LIN to models that use this measure.

The second measure I use is a recursively estimated Hodrick-Prescott filter. This filter has been widely criticised as being atheoretical and somewhat ad hoc, but still remains a heavily used technique of estimating trend output. The measure of the output gap here is simply the difference between log GDP and it's HP trend. One often cited problem with the filter is that it is not very good at the end points of the sample. In fact the 'filter' behaves like a smoother in the middle of the sample (using information from both sides of the observation) and as a true 'filter' at the end of the data, that is it clearly can only use information from one side of the last observation. Many of the problems and issues involved with the HP filter are discussed in Barrel and Sefton (1995) and more recently by Van Norden[11]. There's no reason to throw away information at the beginning of the sample with the HP-filter so I estimate it recursively. For convenience, I retain the previous estimates of the trend at each forecast point. This means that in 98:1 for example the estimates of trend for 90:1 were the ones estimated in 90:1, not the estimate I could now make using information from beyond 90:1. I attach the prefix HP to the models that use the HP measure.

The final measure I use is less mechanical, and not strictly speaking an output gap measure per se. I have used the differenced capacity utilization series for the manufacturing sector (obtained from Datastream). Capacity utilization has long been used as a leading indicator for inflation, see for example Corrado and Mattey ([7] 1997). The same sort of arguments for a nonlinear relationship between capacity utilization and the output gap hold. In fact Razzak ([24] 1998) has used precisely this measure in a study of nonlinearity in the output-gap/inflation relationship in New Zealand. These models are denoted via a CAP prefix.

Plot 1 shows the three measures. Clearly the general shape of the measures is very similar. For most of the time all the measures would lead to the same inference regarding whether the economy is below or above trend.

6 Estimated Forecasting Model Details

The models are referred to in the form 'gap measure, LIN, HP or CAP'-'model type LINear or THReshold', e.g. LIN-LIN is a linear model which uses the rolling linear gap measure. For the LIN and the HP gap, the results were quite similar. In both cases, for each linear forecasting model selected, an output gap term was selected via the SIC[12]. So for these output gap measures the two threshold models estimated

[11] The Bank of Canada web site has a large number of papers dealing with the issue of estimating of estimating output gaps, see http://www.bank-banque-canada.ca/english/wp(y).htm

[12] Remember that there is no requirement that the linear model has to include an output gap term. In fact for the capacity gap model results we will find that the output gap term is not

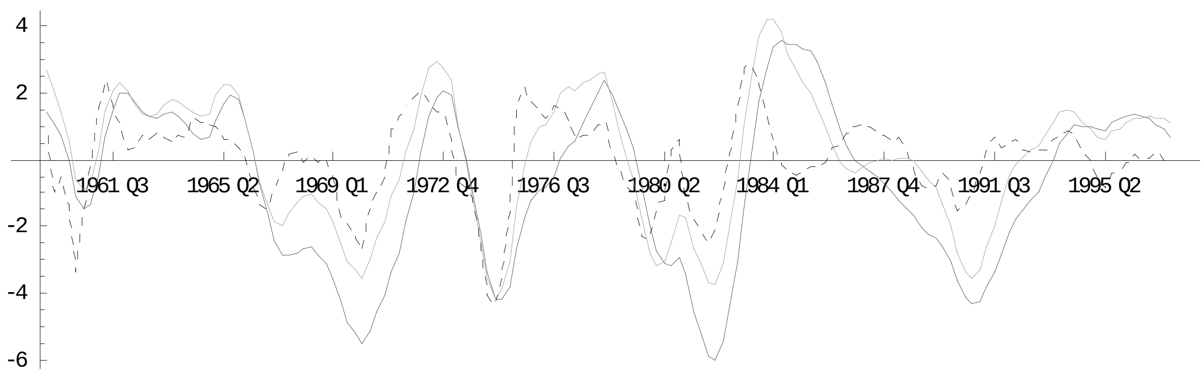

Figure 1: Output Gaps (Dark line - Rolling Linear, Light Line - Recursive HP Filter, Dashed line - Capacity Utilization)

(each period) including the inflation lags selected via the SIC on the linear model, and the respective lag of the output gap. For the LIN measure $ygap_{t-2}was$ chosen, for the HP gap $ygap_{t-3}$ was chosen. Also for both models only π_{t-1} and π_{t-3} were chosen from the set of inflation lags available. For the CAP gap measure, none of the forecasting models included an output gap term, and so a simple model containing π_{t-1} and π_{t-3} was recursively estimated throughout the forecasting exercise. Therefore my automatic procedure for specifying a threshold model would not be applicable. However close inspection of the individual model estimates revealed that the CAP measure was significant throughout the forecasting set, though obviously not significant enough once penalized by the SIC. So, out of interest I did go on and estimate a threshold model with a single lagged (the first lag as this was the most significant in a comparison of the four models which contained a single lag of *ygap* and the two inflation lags) *ygap* term. In fact, having done this the capacity gap measure was still significant in the threshold model, but as will be seen in the next section, this did not help the forecasting performance of the model.

Both the HP and the LIN threshold models displayed similar results. The estimated nonlinearity actually pointed towards a form on mild concavity, consistent with the results of Filardo (1998 [12]). Only when the threshold variable (the lagged output gap term) fell below a critical level (roughly 2 percent below trend[13]) did the model suggest a significantly positive effect from the *ygap* term. At other times, that is when the output gap term was above -2%, there was no significant effect from the output gap term. Tables 1 and 2 show some statistics on the estimated model parameters. Significant (at 95%) parameters are in bold. The estimates were very stable over time as can be seen by looking at the minimum, maximum and mean

selected, despite being significant within the equation.

[13]The estimated threshold for the LIN model was -2.315 until 92Q1 and then -2.562 until 98:2. For the HP model the threshold was 1.413 until 92Q2 then -1.863 until the end of the sample.

	π_{t-1}	π_{t-3}	$ygap_{t-2}$	
LIN-LIN				
Min	**.701**	**.269**	**.024**	
Mean	**.717**	**.291**	**.026**	
Max	**.736**	**.307**	**.031**	
LIN-THR	.		$ygap_{t-2}(upper)$	$ygap_{t-2}(lower)$
Min	**.705**	**.298**	-0.028	**.173**
Mean	**.718**	**.316**	-0.009	**.182**
Max	**.738**	**.326**	0.006	**.194**

Table 1: Forecasting Model Parameter Estimates - LIN

	π_{t-1}	π_{t-3}	$ygap_{t-3}$	
HP-LIN				
Min	**.677**	**.280**	**.132**	
Mean	**.693**	**.301**	**.138**	
Max	**.712**	**.317**	**.149**	
HP-THR			$ygap_{t-3}(upper)$	$ygap_{t-3}(lower)$
Min	**.676**	**.303**	.013	**.247**
Mean	**.693**	**.327**	.030	**.258**
Max	**.726**	**.347**	.049	**.271**

Table 2: Forecasting Model Parameter Estimates - HP

parameter estimates over the 52 period out-of-sample set. The table shows that estimating the linear model gives an averaged out estimate of the effect of the output gap term, as the estimate is between the lower regime and the upper regime estimate of the threshold model. The significance of the lower regime output-gap terms was also considerably stronger than the significance of the output gap term in the linear model.

The effect of the nonlinearity can be seen by plotting the time-varying parameter on the output gap term throughout the out-of-sample forecasting set. Plots 2 and 3 show this parameter over time, and also the constant paramater from the linear model. What we can see is that for much of the sample the effect of the output gap is virtually zero, but for a period in the middle of the sample, the parameter becomes positive.

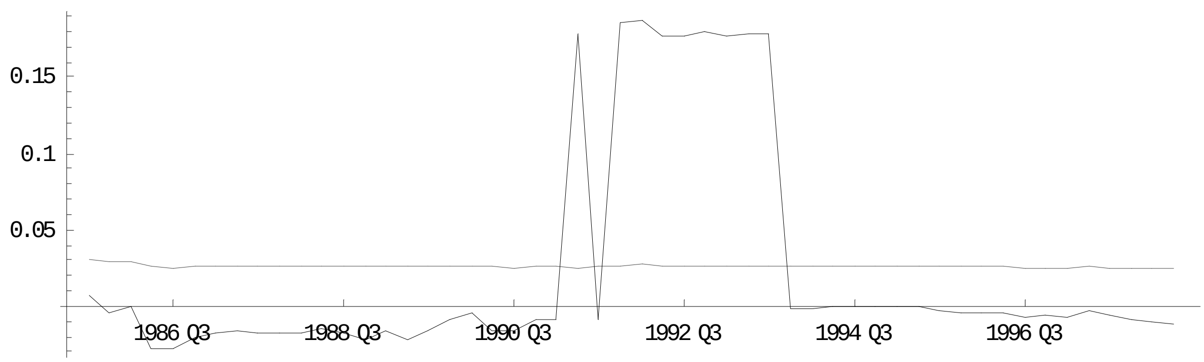

Figure 2: Time Varying Coefficient on the Output Gap Measure when *ygap* is a rolling linear trend

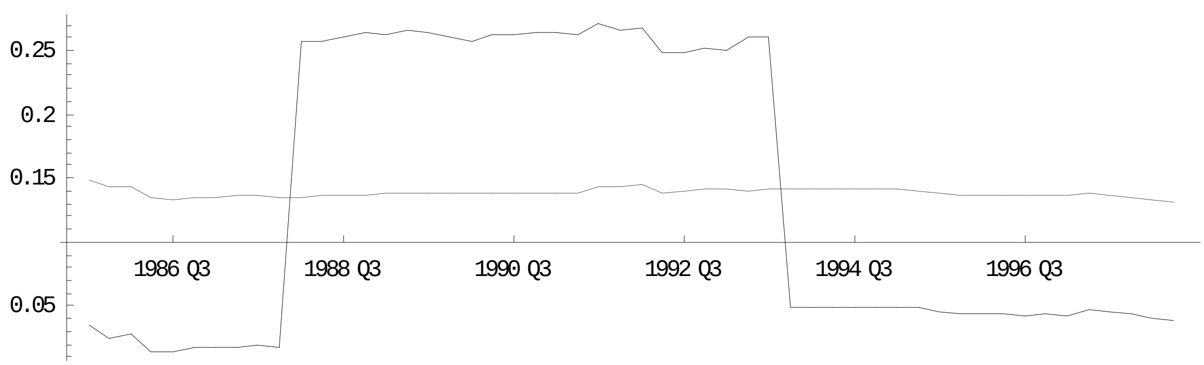

Figure 3: Time Varying Coefficient on the Output Gap Measure when *ygap* is a recursive Hodrick-Prescott filtered estimate

7 Out-of-Sample Forecasting Results

7.1 Forecast Criterion

I have used three out-of-sample forecast metrics. The first two are fairly standard metrics, mean square error (mse), mean absolute deviation (mad). The third metric is the confusion rate, which gives us a measure of ability of the model to forecast the direction of inflation. This last measure has been discussed by Pesaran and Timmerman (1994 [23]) and Swanson and White (1995 [27]). In our case we consider a 2 by 2 matrix, with the row corresponding to predictions of up or down, and the columns corresponding actual ups or downs. The sum of the off-diagonals divided by the total number of observations out-of-sample represents the number of times the model forecasted the wrong direction. I report this statistic, and also the χ^2 test of independence p-values, based on the null hypothesis that the model is of no value in predicting the direction of the inflation rate. χ^2 is calculated simply as $\chi^2 = \sum_{i=1}^{k}(f_{oi} - f_{ei})^2/f_{ei}$, where i ranges over the confusion matrix entries and f_{oi} are the observed numbers in the cells, and f_{ei} are the expected number. I also report $\phi = \sqrt{\chi^2/T}$. This is another measure of the directional predictability and in effect gives the degree of diagonal concentration, ranging from 0 when the actual and predicted directions are completely independent to 1 when they are perfectly matched.

The second main way in which I will compare the forecastability of different models is via a number of pair-wise comparison tests. I use the predictability comparison method introduced by Diebold and Mariano ([10] 1995) and applied in the inflation forecasting context by Chen and Swanson ([4] 1997). The 'loss-differential' test requires a sample of differentials, $\{d_t\}_{t=1}^{T}$. In the generic case, given some loss metric, $g()$, we create the d_t series as $d_t = g(z_{t,1}) - g(z_{t,2})$. For our two metrics mse and mad, the $g()$ functions are $g(x) = x^2$ and $g(x) = |x|$ respectively. We then use this series to estimate the large sample statistic

$$D = \overline{d}/[T^{-1}2\pi\widehat{f_d}(0)] \sim N(0,1)$$

$\widehat{f_d}(0)$ is an estimate of the spectral density of the loss differential at frequency 0. We follow Diebold and Mariano and use obtain a consistent estimate by calculating a (two-sided) weighed sum of the available sample autocovariances, specifically

$$2\pi\widehat{f_d}(0) = \sum_{\tau=-(T-1)}^{(T-1)} L(\frac{\tau}{S(T)})\widehat{\gamma}_d(\tau)$$

and

$$\widehat{\gamma}_d(\tau) = \frac{1}{T}\sum_{t=|\tau|+1}^{T}(d_t - \overline{d})(d_{t-|\tau|} - \overline{d})$$

	mse	**mad**	**CR**	**p-value**	ϕ
RW	0.674	0.635			
AR(π)	0.512	0.582	0.353	0.069	0.255
LIN-LIN	0.544	0.577	0.333	0.031	0.302
LIN-THR	0.507	0.580	**0.314**	0.015	0.339
HP-LIN	0.518	0.579	0.353	0.067	0.257
HP-THR	**0.497**	**0.568**	**0.314**	0.015	0.339
CAP-LIN	0.503	0.577	0.373	0.112	0.223
CAP-THR	0.538	0.595	0.392	0.138	0.208

Table 3: Forecasting Results

$L(\frac{\tau}{S(T)})$ is the lag window, and we use Diebold and Mariano's suggestion to use a uniform lag window i.e.

$$\begin{aligned} L(\frac{\tau}{S(T)}) &= 1 \text{ for } \left|\frac{\tau}{S(T)}\right| \leq 1 \\ &= 0 \text{ otherwise} \end{aligned}$$

We also set $S(T) = k - 1$, based on the idea discussed by Diebold and Mariano that as optimal $k - ahead$ forecasts imply at most $k - 1$ dependence, then the lag window need only include $k - 1$ autocovariances as others should be zero. This asymptotic test method is robust to non-zero mean errors, non-normality, and contemporaneous correlation. It's easily applicable to arbitrary $g()$ functions, and is easily implementable.

7.2 Forecast Results

Table 3 contains the forecasting results. The best results for each metric are in bold. In the table I also refer to two other models. The fist RW is a simple random walk with no drift. This clearly cannot give directional forecasts, but is a standard benchmark model for mse and mad metrics. The second model (AR(π)) is a linear model that contains two lags of inflation only, π_{t-1} and π_{t-3}. These were the two lags selected by the SIC, and this model would be the optimal model if there were no output gap terms permissible in the equation[14]. This model actually performs rather well, it beats the linear output-gap models in terms of mse and mad, but is not much better in terms of directional forecastability. The RW model does not perform well in comparison to the other models as might be expected, but I think it would be misleading to justify the usefulness of the output gap terms by comparing the output-gap models (LIN-LIN and HP-LIN) with this model, a better comparison is with the AR(π) model.

[14]It also happened to be the best model selected by SIC when I allowed a capacity gap measure to enter the equation.

	RW	AR(π)	LIN-LIN	LIN-THR	HP-LIN	HP-THR	CAP-LIN	CAP-THR
RW	0	2.257	1.735	1.668	1.527	2.044	2.128	1.11
AR(π)	−2.257	0	0.1998	0.1102	0.1022	0.7054	0.663	−0.6147
LIN-LIN	−1.735	−0.1998	0	−0.1652	−0.1791	0.6583	−0.001	−0.5854
LIN-THR	−1.668	−0.1102	**0.1652**	0	0.04462	0.8422	0.1412	−0.525
HP-LIN	−1.527	−0.1022	0.1791	−0.04462	0	1.883	0.0779	−0.508
HP-THR	−2.044	−0.7054	−0.6583	−0.8422	**−1.883**	0	−0.4351	−0.9988
CAP-LIN	−2.128	−0.663	0.001	−0.1412	−0.0779	0.4351	0	−1.19
CAP-THR	−1.11	0.6147	0.5854	0.525	0.508	0.9988	1.19	0

Table 4: MAD Pairwise tests

	RW	AR(π)	LIN-LIN	LIN-THR	HP-LIN	HP-THR	CAP-LIN	CAP-THR
RW	0	4.292	2.344	2.885	2.761	2.799	3.603	2.015
AR(π)	−4.292	0	−0.7743	0.1525	−0.1645	0.4159	0.6984	−0.7061
LIN-LIN	−2.344	0.7743	0	1.126	1.248	1.263	1.059	0.1245
LIN-THR	−2.885	−0.1525	**−1.126**	0	−0.4314	0.53	0.1902	−0.9236
HP-LIN	−2.761	0.1645	−1.248	0.4314	0	0.9132	0.4808	−0.5005
HP-THR	−2.799	−0.4159	−1.263	−0.53	**−0.9132**	0	−0.1906	−1.338
CAP-LIN	−3.603	−0.6984	−1.059	−0.1902	−0.4808	0.1906	0	−1.397
CAP-THR	−2.015	0.7061	−0.1245	0.9236	0.5005	1.338	1.397	0

Table 5: MSE Pairwise Tests

Whilst the LIN-LIN model does beat the RW model convincingly, it is not quite so clear when compared with the AR(π) model, where the latter model actually beats LIN-LIN by mse[15].

The best model for each criterion is a nonlinear model. The HP and the LIN threshold models beat both the RW and the AR(π) benchmark models, and also with the exception of mad for the LIN model, they beat their linear counterparts. The capacity measure based models do not fare well in comparison with the other models. The CAP-LIN model is better than the AR(π) model in terms of mse and mad, but worse in terms of direction. The CAP-THR model, despite having significant parameters, fared worse of all. In a sense there is some encouragement to be had from this fact because the SIC did not suggest going forward and augmenting the linear model with the nonlinear part. Following the SIC would have been justified therefore in this case.

Tables 5 and 4 contain the results of the pairwise tests. In bold I have marked out a particular comparison of interest, which is the nonlinear versus the linear model. We

[15] Chen and Swanson (1997) argued the linear model with an output gap term was significantly better than a RW model, but the point is this wouldn't justify a claim that the inclusion of the output gap term significantly improves the forecasting ability of a model without the output gap term.

can see that only when we compare HP-THR against HP-LIN do we find a significant improvement, but this is not to be sniffed at. Significant improvements are not common in these exercises. Taken together with the fact that HP-THR model has the lowest confusion rate overall, and also considerably lower than it's linear counterpart (HP-LIN) there is convincing evidence that a nonlinear model with a HP output-gap is a good model for forecasting. We can also see that all the models bar the CAP-THR for mad, do significantly better than the RW. So there is some robustness here, using this simple SIC method, followed by the threshold augmentation, we are not likely to end up with very bad forecasts, unlike some of the other results discussed earlier which used very complex nonlinear augmentations.

8 Discussion of Results

The simple threshold models estimated here appear to be very attractive from a forecasting perspective. A nonlinear model performs best for all three forecast criterion, and there are some significant comparisons to be found as well. However there is a caveat. The nonlinearity has been found to be useful, but only within this set of explanatory variables, and only compared to simply specified benchmark models. Firstly, it could be the case that were we to model expectations more precisely (by introducing forward looking measures for example) or to include more lags of inflation, the case for nonlinearity may be weakened. In the actual generation of the inflation rate a number of observable variables may contribute linearly, but given we haven't included all the relevant variables, we are left with the optimal specification (as opposed to the 'true' specification) being nonlinear. The nonlinear specification is optimal given the set of explanatory variables under consideration. Also the estimation of the output-gap term is not necessarily an innocent procedure. It might be that some 'true' measure of the output gap would enter the equation linearly, but the filtering methods I used here have induced some nonlinearity into the equation. So we could not claim on the strength of these results that the output-gap drives inflation in a nonlinear way, nor would we. However it is fair to say that given these estimates of the output gap and this set of explanatory variables, a nonlinear model is more appropriate for modelling the relationships than a linear model. Of course this sort of statement will be more relevant than not in econometrics, where we sometimes have very weak theory as to the set of relevant regressors.

9 Conclusions

I have examined the usefulness of following a very simple model estimation and specification procedure to the task of forecasting quarterly inflation one-step-ahead. Whilst the nonlinear augmentation procedure may not be optimal, it appears to be quite robust, and undoubtedly provides a pay-off in terms of improved forecastability. However it would be worthwhile examining other procedures for specifying the nonlinear model. Given the two-regime threshold functional form I use here, I could treat the specification of the threshold model completely independently from the linear specification. One obvious way to do this would be to 'SIC' the nonlinear model as well, searching over all possible nonlinear specifications just as I do for the linear model. This is because just because the first lag (say) of the output gap term is included in the linear model, it does not follow that the first lag of the output gap term ought to appear in the optimal threshold model faced with the same set of potential regressors. To what extent these results are significant in a wider sense, when compared to other benchmark models is also of interest. In another paper I use some of these forecast results and compare them to survey measures, the results there support these forecasts as being useful. Stockton and Glassman (1987) also found that simple Phillips curve based models were not consistently beaten by other econometric model using other variables. Chen and Swanson (1997) also looked at models (linear and nonlinear) which had the potential to include a much wider range of variables, including excess labour supply, money supply growth and interest rates. Interestingly their 'best' models only used lagged inflation and lagged output gap terms, so this evidence would further support the fact the forecasts presented here would not be easily beaten.

Finally it will be of interest to tie up the methods used here more closely to the current literature on policy analysis of nonlinear Phillips curves. My model suggests a mild concavity, at low levels of the output gap there is a kink from a positive slope to an insignificant slope. This type of kink was also found by Filardo (1998 [12]) over a similar sample. However, he also found another regime at higher levels of the output gap. In this regime, which represents an overheated economy, he found another set of circumstances when the output-gap exerts significant pressure on inflation. Clearly it would be of interest to see how a three-regime model would compare with the two-regime model explored in this paper.

References

[1] R. Barrell and J. Sefton (1995), *Output gaps, some evidence from the UK, france and germany*, National Institute Economic Review (1995).

[2] N. Cameron and S. Moshiri (1997), *Neural network vs. econometric models in forecasting inflation*, Tech. report, University of Manitoba, Working Paper.

[3] C. Chen, R. McCulloch, and R. Tsay (1996), *A unified approach to estimating and modelling linear and nonlinear time series*, Tech. report, University of Chicago, Graduate School of Business, 1996.

[4] X. Chen and N. Swanson (1997), *Semiparametric ARX neural network models with an application to forecasting inflation*, Tech. report, Penn State, 1997.

[5] P. Clark, D. Laxton, and D. Rose (1996), *Asymmetry in the u.s. output-inflation nexus*, IMF Staff Papers **43** (1996), 216–251.

[6] P. Clark and D. Laxton (1997), *Phillips curves, phillips lines and the unemployment costs of overheating*, IMF Working Paper (1997).

[7] C. Corrado and J. Mattey (1997), *Capacity utilization*, Journal of Economic Perspectives **11** (1997), 151–167.

[8] G. Debelle and D. Laxton (1997), *Is the phillips curve really a curve? some evidence for canada, the united kingdom and the united states*, IMF Staff Papers (1997).

[9] M. Deutsch, T. Terasvirta, and C. Granger (1994), *The combination of forecasts using changing weights*, International Journal of Forecasting **10**, 47–57.

[10] F.X. Diebold and R.S. Mariano (1991), *Comparing predictive accuracy i: An asymptotic test*, Tech. Report 52, Federal Reserve Bank of Minneapolis.

[11] C. Dupasquier and N. Ricketts (1998), *Non-linearities in the output-inflation relationship: Some empirical results for canada*, Tech. Report 98-14, Bank of Canada, 1998.

[12] A. Filardo (1994), *Business-cycle phases and their transitional dynamics*, Journal of Business and Economic Statistics **12** (1994), 299–308.

[13] R. Gordon (1997), *The time-varying NAIRU and its implicatins for economic policy*, The Journal of Economic Perspectives **11** (1997), 11–32.

[14] B. Hansen (1996), *Testing for structural change in conditional models and the bootstrap*, Tech. report, Boston College.

[15] B. Hansen (1997), *Inference in TAR models*, Studies in Nonlinear Dynamics and Econometrics **2**, no. 1.

[16] R. Hillman (1998), *Mixtures of forecast combinations*, Tech. report, Financial Econometrics Research Center, manuscript.

[17] R.J.T. Hillman (1998), *Econometric modelling of nonlinearity and nonstationarity in the foreign exchange market*, Ph.D. thesis, University of Southampton, 1998.

[18] D. Laxton, D. Rose, and D. Tambakis (1997), *The u.s. phillips curve: The case for asymmetry*, Tech. report, IMF Working Paper, 1997.

[19] T. Lee, H. White, and C. Granger (1993), *Testing for neglected nonlinearity in time series models*, Journal of Econometrics **56** (1993), 269–290.

[20] R. Lucas (1972), *Expectations and the neutrality of money*, Journal of Economic Theory **4** (1972), 103–24.

[21] S. Moshiri, N. Cameron, and D. Scuse (1997), *Static, dynamic, and hybrid neural networks in forecasting inflation*, Tech. report, University of Manitoba, Working Paper, 1997.

[22] D. Peel and A. Speight (1998), *The nonlinear time series properties of unemployment rates: Some further evidence*, Applied Economics **30**, 287–294.

[23] H. Pesaran and A. Timmermann (1992), *A simple non-parametric test of predictive performance*, Journal of Business and Economic Statistics **10**, 461–465.

[24] W. Razzak (1998), *A policy implication of asymmetric phillips curve in a small open economy*, Tech. report, Reserve Bank of New Zealand, 1998.

[25] J. Stiglitz (1997), *Reflections on the natural rate hypothesis*, The Journal of Economic Perspectives **11** (1997), 3–10.

[26] D. Stockton and J. Glassman (1987), *An evaluation of the forecast performance of alternative models of inflation*, The Review of Economics and Statistics (1987), 108–117.

[27] N. Swanson and H. White (1995), *Forecasting economic time series using adaptive versus non-adaptive and linear versus nonlinear econometric models*, Tech. report, Pennsylvania State University, 1995.

[28] H. White (1989), *Learning in artificial neural networks: A statistical perspective*, Neural Computation, 425–464.

[29] A. Yates (1998), *Downward nominal rigidities and monetary policy*, Bank of England Working Paper (1998).