Review by Tony Yates of: 'Inflation/unemployment regimes and the instability of the Philips Curve', by Ormerod, Rosewell and Phelps

Substantive comments

I think this paper has chosen an important and topical subject. There has in recent years been a resurgence in interest in using sophisticated, theory-neutral statistical tools to characterise the dynamics of inflation and unemployment. In other literatures these papers have served to sharpen the set of stylised facts that theoretical models need to answer.

I had many minor comments, which are set out below in the order in which they were prompted by the text of the paper. I would summarise my substantive response to the paper in the following general remarks.

First, it was, in the end, not that clear what the authors wanted us to take away from the statistical analysis. For an example of what I was looking for, take the paper most similar to this, by Luca Benati, in the JMCB of 2007. This paper, in a nutshell, says: the correlation between inflation and unemployment varies greatly over time, and theory gives us plenty of reasons why we would expect it to. Are Ormerod and co-authors intending to send the same message, but have us note that this is confirmed by using clustering, rather than Benati's spectral analysis? It's not that surprising that a correlation between two observable macro variables changes over time: so it's not that compelling a finding to report that it does. Modern dynamic stochastic general equilibrium models with sticky wages, for example, would tell you that this correlation would depend on the mix of shocks hitting the economy; the nature of the monetary and fiscal policy regime – what stabilisation policy they are engaged in; whether there are changes in labour market institutions, or other features of economic structure that would change model parameters.

Second, I think the paper should address itself more closely to its antecedents (e.g. the Benati paper cited above, and others), explaining what different insights can be got using clustering, but using the prior studies as a starting point. The alternative methods I would list as: stochastic time-varying parameter models, popularised by Cogley and Sargent and Sims; deterministic time-varying coefficient models, e.g. the work of George Kapetanios; rolling regression or spectral analysis. What are its costs and benefits? Are they dependent on the context? If so, are they particularly useful here? Are there limiting cases in which clustering converges to one of the other methods?

Third, the statistical technique being tested here seeks to find discrete numbers of values for the inflation-unemployment correlation. But most modern macro theories would not predict that there should be a discrete number of such correlations. They would predict, as already emphasised above, that the correlation should depend on the mix of shocks over time, which we would expect to change in a continuous fashion, and not switch between small numbers of values; and also on evolutions in the monetary and fiscal policy regime, and in other institutions. Is there a sensible theory that would predict only small numbers of values for the inflation-unemployment correlation? If not, is clustering not starting out with a model that we have good theoretical grounds for rejecting at the outset?

Minor comments arising from the draft

I don't think it's fair to say that economic theory offers no guidance as to how long is the long run. Friedman and Phelps who first formalised the notion of the long run Phillips Curve both had in mind that the long run was the length of time it took for prices and wages to adjust. Modern 'New Keynesian' models embody a variety of frictions including price, wage stickiness, and sluggishness in consumption and investment – all these make quite precise predictions about how long it takes for the real effects of a monetary shock to dissipate, or, in the language of this paper, over what horizon the Phillips Curve is vertical.

Page 3 – what is meant by the term 'the level of the Phillips Curve'? The intercept?

There is quite a lot of empirical work on instability in the Phillips curve, which this paper should address and discuss in order to establish the motivation for another piece. Two examples: Luca Benati's JMCB piece; and work by Jan Groen of the New York Fed. Benati looks at estimates of the :Philips correlation – just as here – and how it has changed over time, using spectral analysis. Jan Groen's piece estimates changes in the Phillips Curve in a structural model.

Page 8 – can you define what is meant by an 'attribute' of a data point? At the bottom it is written that fuzzy clustering 'contains more information in its output' than classical clustering. What does this statement mean?

Page 10. We are introduced to the terms 'classical' and 'fuzzy'. If I understand correctly, 'classical' is not meant to invoke the opposite of 'Bayesian'; and 'fuzzy' here has no connection to 'fuzzy logic'. If so would some comment to this effect be helpful?

Discussion of the clustering methodology: why use this technique? What advantages and disadvantages does it have from others, e.g. using models of stochastic parameter change – see, for example, the work of Cogley and Sargent in the NBER macro annual discussing changes in inflation dynamics – or rolling regression or rolling spectral analysis, or models of deterministic structural change (see, for example, the work of George Kapetanios at QMC)?

I find the exposition of the clustering methodology difficult to follow. I would prefer it to be less disembodied from the example at hand. As I understand it, the technique seeks to discover statistical clusters in the inflation-unemployment correlation. The technique allows us to estimate the probability that an individual observation was drawn from one or other of the clusters. If this is right, then perhaps that could be stated briefly, up front.

I don't follow the exposition of the clustering algorithm. Suppose we define at the outset a cluster number equal to the number of observations. Doesn't this straight away produce the lowest value of the objective function? Presumably there is some penalty for having a larger number of clusters. Following this thought, I don't understand how the use of the Dunn coefficient is built into the use of the clustering

algorithm. Presumably the procedure has the following form: follow algorithm, check value of Dunn coefficient. If value unsatisfactory, go back and (and what? Start with a new initial condition? Trim the number of clusters by one, or what?). There is an ominous comment about the use of judgement in using the Dunn coefficient. Could the authors be explicit about what judgement is being used in this paper, and to what end?

Why are the initial conditions randomised? Why not start with the smallest number of clusters and build up? Using an analogy from other econometric techniques isn't the idea to try to get away with as small a number of parameters as possible, to maximise the efficiency with which they are estimated? Does this apply here? If not, why not?

Page 12: at the top of this page the text talks about '100 iterations of the test'. What is being referred to here?? Is this 100 runs of the algorithm with 100 different initial conditions?

Page 14. I don't understand precisely what the transition probabilities are that are being calculated at the bottom of page 14. An equation might help.