

Do Markov-Switching Models Capture Nonlinearities in the Data? Tests using Nonparametric Methods

Robert V. Breunig

Centre for Economic Policy
Research, Research School of Social
Sciences and School of Economics,
The Australian National University,
Canberra ACT 0200 Australia

Adrian R. Pagan

Research School of Social Sciences,
The Australian National University,
Canberra ACT 0200, Australia
and Oxford University, Oxford,
United Kingdom

Abstract: Markov-switching models have become popular alternatives to linear autoregressive models. Many papers which estimate nonlinear models make little attempt to demonstrate whether the non-linearities they capture are of interest or if the models differ substantially from the linear option. By simulating the models and nonparametrically estimating functions of the simulated data, we can evaluate if and how the nonlinear and linear models differ.

Keywords: Markov-switching models; Nonparametric estimation; Simulation methods

1. INTRODUCTION

The purpose of this paper is to provide an example of how simulation methods, in combination with non-parametric density and regression estimation, may be used to evaluate the success (or failure) of non-linear models in capturing important characteristics of the data. As an example, we look at Markov-switching (MS) models, popularized in the economics literature by [7]. MS models have become a popular alternative to linear autoregressive models in part due to their intuitive appeal, in that they allow a latent state process to cause “switching” between different linear regimes. This corresponds to the popular conception of underlying states of the economy, for example high-growth or low-growth states in GDP and high-volatility and low-volatility states in the stock market.

[9] provided an early example of the use of simulation and nonparametric techniques applied to MS models while [11] and [4] apply nonparametric conditional mean estimation to a selection of non-linear models. The idea of comparing nonparametric estimates of densities and models goes back to the late 1980s at least, and was used by [10] and [1] in analyzing financial data.

In this paper we consider a simple Markov-switching model estimated by Bodman in [2] using Australian unemployment data. The example in this paper is illustrative of the problems which often arise in estimating and using Markov-switching models. Basically these are failures of the algorithm to reach the global optimum of a likelihood and an inability to fit important features of the data. Moreover, the fitted MS model is often indistinguishable from simpler

alternatives. [3] explore a number of other examples of MS models which share the same problems.

2. MARKOV-SWITCHING MODELS

Consider the following example of a Markov-switching model which describes the movement of some stationary economic variable x_t as a function of its p most recent lagged values and a binary state variable, s_t , taking values of 0 or 1:

$$x_t = \mu_0(1 - s_t) + \mu_1 s_t + \sum_{l=1}^p \phi_l (x_t - \mu_0(1 - s_t) - \mu_1 s_t) + \sigma v_t \quad (1)$$

s_t evolves as a first-order Markov process with probability P_{00} of remaining in state 0 next period conditional on being in state 0 this period while P_{11} is defined analogously. Extensions of this simple model have been proposed: using more than two states; allowing dependence of the error variance, σ , and the autoregressive parameters, ϕ_l upon the state; allowing the state variable to evolve as a higher-order Markov process; and allowing the transition probabilities to depend upon the duration of time spent in a given state. The model (1) may be estimated by Maximum Likelihood using an algorithm proposed by Hamilton in [7]. If $\mu_0 = \mu_1$, the model collapses back to a linear autoregressive model; otherwise it is a non-linear model.

Our argument is that it should be incumbent upon researchers who propose this non-linear alternative to show (a) that the linear model, i.e. the case where $\mu_0 = \mu_1$ and there is only a single state, can be rejected; (b) that the MS model

used for the tests in (a) is an acceptable specification; and (c) that the non-linear model captures important features of the data better than the linear model. There is an extensive literature on the first of these requirements e.g. [5] and [8] have provided tests of the MS model against a one-state null hypothesis. The second has received much less attention. [3] propose a combination of moment-based and informal tests to satisfy this requirement and show how they can be used to test for general misspecification in MS models. In fact these tests can be interpreted in terms of the non-parametric methods that we will use in this paper. Finally, although one can view the tests carried out in (a) as tests of non-linearity, it is often difficult to know what governs the outcome of those formal tests and what is the extent of non-linearity i.e. to gain an appreciation of whether the degree of non-linearity is important for the uses to which the model is to be put.

3. INFORMAL TESTS

The central idea of this paper is to simulate from the proposed model taking the parameter estimates, $\hat{\theta}$, as given. Using that simulated data, we can use nonparametric techniques to compute quantities such as $f_M(x_t | \hat{\theta})$, the implied density of x_t from the model, and $E_M(x_t | x_{t-j}; \hat{\theta})$, the implied conditional mean from the model, and compare these to the respective quantities from the data. In order to minimize the degree of error in estimating these functions, we take a very large number of simulated draws (60,000 in the application below.) Comparisons are made using simple graphs, providing a very powerful tool for quickly identifying problems with the models and the ways in which the non-linear and

linear models differ. The visual check of the “fit” between the data and the implied quantity from the model gives the informal test.

The intended use of the model should inform the choice of functions to be examined. In the empirical example below we consider two aspects of the data: the empirical distribution of quarterly changes in the Australian unemployment rate, x_t , and the mean of the unemployment rate changes conditional on their one-period lagged values. Density comparisons will be particularly important where one would like to compare MS and alternative models on the basis of characteristics of the density function, an example being in models of financial data where a criticism of commonly used models such as GARCH is their failure to capture the concentration of returns around zero, see [10]. The latter choice will be important for forecasting applications since then it is the conditional mean that is the relevant quantity.

3.1 Nonparametric Estimation

Once we have generated a large number of simulated draws, n , from the model, we estimate the density function at a point z using the non-parametric kernel density estimate introduced by [14] and [13].

$$\hat{f}_M(z) = \frac{1}{nh} \sum_{s=1}^n K\left(\frac{x_s - z}{h}\right) \quad (2)$$

where $K(\cdot)$ is a smooth kernel function, h is a window width which controls the smoothness of the estimate and x_s are simulated observations from the fitted MS model. We can also find an estimate of this density using the same formula but with x_s replaced by the data x_t and the summation running from $t=1, \dots, T$. A different window width to that used with the simulated output will also be

adopted. [12] provide a general survey of nonparametric estimation, including discussion of choice of kernel and window width for a wide range of problems. We estimate the two densities using a normal kernel and a range of values for z ; in practice we use as z the T data points $x_1 \dots x_T$ in the data set. This allows us to create a graphical comparison of the density function estimated from the data and from the model at the chosen points z . In the density estimates presented below, we choose the bandwidth, h , for the actual data by leave-one-out cross-validation whereas for the simulated data, we use $h = 1.06\sigma n^{-1/5}$ as a simple plug-in bandwidth. With 60000 observations, the technique is not sensitive to bandwidth selection.

Another quantity computed is a conditional mean at the chosen values, z , for the conditioning variable. Making use of the simulated values, we estimate the mean of x_t conditional upon values lagged one period using the Nadaraya-Watson kernel regression

$$\hat{E}_M(x_t | x_{t-1} = z; \hat{\theta}) = \frac{\sum_{s=2}^n x_s K\left(\frac{x_{s-1} - z}{h}\right)}{\sum_{s=2}^n K\left(\frac{x_{s-1} - z}{h}\right)} \quad (3)$$

We use a normal kernel and the robust bandwidth suggested in [15]: $h = .9 * \min(IQR/1.34, \sigma)$ where IQR is the inter-quartile range. However, unlike the density comparison, we don't compute a conditional mean from the data. Instead we compare the conditional mean implied by the model with a cross plot of the data against its lagged value.

4. APPLICATION

Bodman, in [2], fitted the two state MS model (1) to the first differences of the quarterly Australian unemployment rate, x_t , for the time period 1959:3 to 1997:3. The second column of Table 1 provides his estimates of the parameters.

Using these parameter estimates, we generated 63000 observations from the implied model. We drop the first 3000 observations to remove the effect of starting values. Consequently, estimates of $f_M(x_t | \hat{\theta})$ and $E_M(x_t | x_{t-j}; \hat{\theta})$ are based upon sample sizes of 60,000. Figure 1 presents the nonparametric density estimates from the data, the linear AR(4) model, and the MS model.

[Figure 1 about here]

Clearly the linear model fails to capture certain key aspects of the data including the high peak just below zero and the rapid decline in the density between zero and one-half. However, the degree to which the MS model fails to capture the distribution of the data is shocking. Here we are making a simple informal visual comparison. Since the densities are so clearly different, no formal test seems necessary. However, it is possible to derive formal tests, as done in [1]. [3] derive formal tests based upon testing moments of the data against the simulated moments of the model. They show that failure to match the simple moments such as the mean and variance may be used as a check of model convergence.

Many researchers, e.g. p.332 of [6] and Hamilton's web page, report convergence problems when estimating Markov switching models by maximum likelihood. The likelihood function will typically not be globally concave and in cases where

the non-linear specification is imposed on a linear model, the likelihood function will have flat portions.

To investigate the possibility that the failure of Bodman's MS model to match the data, as evidenced by Figure 1, arose from convergence problems we re-estimated the model using the data provided to us by Bodman and a variety of programs, including that on Hamilton's web site. The revised parameter estimates are given in the third column of Table 1.

Table 1. Comparison of Bodman and Alternative Estimates for an MS model.

	Bodman estimate	Revised estimate
μ_0	.486	.6952
μ_1	-.361	-.0399
ϕ_1	.022	.4174
ϕ_2	.147	.1163
ϕ_0	-.079	.1664
ϕ_0	-.185	-.3273
σ	.326	.2148
P_{00}	.83	.48
P_{11}	.89	.927

Figure 2 shows that the model based upon these revised estimates implies a distribution much closer to the data density and superior to the linear model. Hence it is clear that any conclusions drawn from Bodman's original model must be treated with caution due to the failure to converge. The revised model does a quite remarkable job of picking up the nearly flat stretch in the right-hand tail of the distribution and matches the modal point of the distribution much better than

the linear model. It does fail slightly to capture the peak of the data, but is clearly a much better match than that of the linear model.

[Figure 2 about here]

Figure 3 presents a cross-plot of the data against its own lagged values and the conditional mean implied by the linear model, Bodman's original estimates, and our revised estimates. The conditional mean of the linear model is not calculated by simulation, but by using the theoretical conditional mean implied by the AR(4) estimates. For the two non-linear MS models, the implied conditional means are calculated by nonparametric regression (see equation (3) above) using the same set of 60,000 simulated observations used to construct Figures 1 and 2. Again, the obvious problems with Bodman's fitted model are apparent in that the implied conditional mean is substantially different from the relation evident in the data. When we simulate Bodman's originally fitted model, we observe a negative relationship between unemployment changes and their lagged values above values of x of .8. This negative relationship is clearly not supported by the cross plot of data points.

It is also clear that there is very little difference between the conditional expectations found from the revised MS model and a linear model, at least in that part of the variable space where there are a reasonable number of data points. Thus, for an exercise such as one-step ahead forecasting, where it is the conditional mean that is used, we would not expect any gains from using the MS model. Since estimating the MS model is more complicated and involves estimating three additional parameters, it seems as if the gain from this extra complexity is rather small. It is

interesting that the revised MS model is attracted to the one very large change (over 1.5%) in the lagged unemployment rate and "fits" this data point exactly.

In this application, whether the nonlinear model withstands scrutiny depends upon one's stance regarding the importance of the density estimate relative to the conditional mean. For problems where the central concern is matching the unconditional distribution of the data, the MS model would appear superior. For conditional mean modelling, however, the more parsimonious linear model has an equal claim to the MS model.

It is worth noting that the comparison being performed here is based upon comparing a linear model with normal errors to an MS model, i.e. it effectively asks if the density function of x_t is normal. This is also the assumption behind standard tests of the number of states in MS models; when there is more than one state the density of x_t will be non-normal. So these tests may lead one to reject the linear model in favor of the MS model in the event that x_t is not normally distributed, even though neither may be a good representation of the data. In this example, the density implied by the MS model does appear to be a good match with the data. However, if forecasting of the quarterly unemployment data is the objective, then, in the choice between the linear model and the MS model, the linear model would appear to be preferable on grounds of simplicity.

5. CONCLUSIONS

The techniques outlined here have the advantage of being easy to apply, even for very complicated models. Clearly the choice of which conditional moments and which densities to examine depend upon the particular application. In the example presented here, the techniques uncovered a convergence problem with the algorithm that was used to estimate the model. Even after correcting the estimates, there is little evidence that the non-linear model performs better than the linear alternative in ways that would be important for most economic applications.

6. REFERENCES

- [1] Ait-Sahalia, Y., Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9 (1996) 385-426.
- [2] Bodman, P., Asymmetry and Duration Dependence in Australian GDP and Unemployment, *Economic Record* 74 (1998) 399-411.
- [3] Breunig, R. and A.R. Pagan, Some Simple Methods for Assessing Markov Switching Models, mimeo, Australian National University (available at <http://econrsss.anu.edu.au/~breunig/>), 2001.
- [4] Clements, M.P. and A.B.C. Galvao, Conditional Mean Functions of Non-Linear Models of US Output, mimeo, University of Warwick, 2000.
- [5] Garcia, R., Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, *International Economic Review* 39(3) (1998) 763-788.
- [6] Goodwin, T.H., Business-Cycle Analysis with a Markov-Switching Model, *Journal of Business and Economic Statistics* 11 (1993) 231-339.
- [7] Hamilton, J.D., A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* 57 (1989) 357-384.

- [8] Hansen, B., The Likelihood Ratio Test Under Non-standard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* 7 (1992) S61-S82.
- [9] Harding, D. and A.R. Pagan, Dissecting the Cycle: A Methodological Investigation, *Journal of Monetary Economics* 49 (2002) 365-381.
- [10] Pagan, A.R., The Econometrics of Financial Markets, *Journal of Empirical Finance* 3 (1996) 15-102, 1996.
- [11] Pagan, A.R., The Getting of Macroeconomic Wisdom, Invited Address to the World Economic Congress of the International Economic Association in Buenos Aires (available at <http://econrsss.anu.edu.au/staff/adrian/index.html>), 1999.
- [12] Pagan, A.R. and A. Ullah, *Nonparametric Econometrics*, (Cambridge University Press, New York, 1999).
- [13] Parzen, E., On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics* 33 (1962) 1065-1076.
- [14] Rosenblatt, M., Remarks on Some Nonparametric Estimates of Density Function, *Annals of Mathematical Statistics* 27 (1956) 832-837.
- [15] Silverman, B., *Density Estimation for Statistics and Data Analysis*, (Chapman and Hall, London, 1986).

Figure 1: Density implied by Bodman MS Model and AR(4)

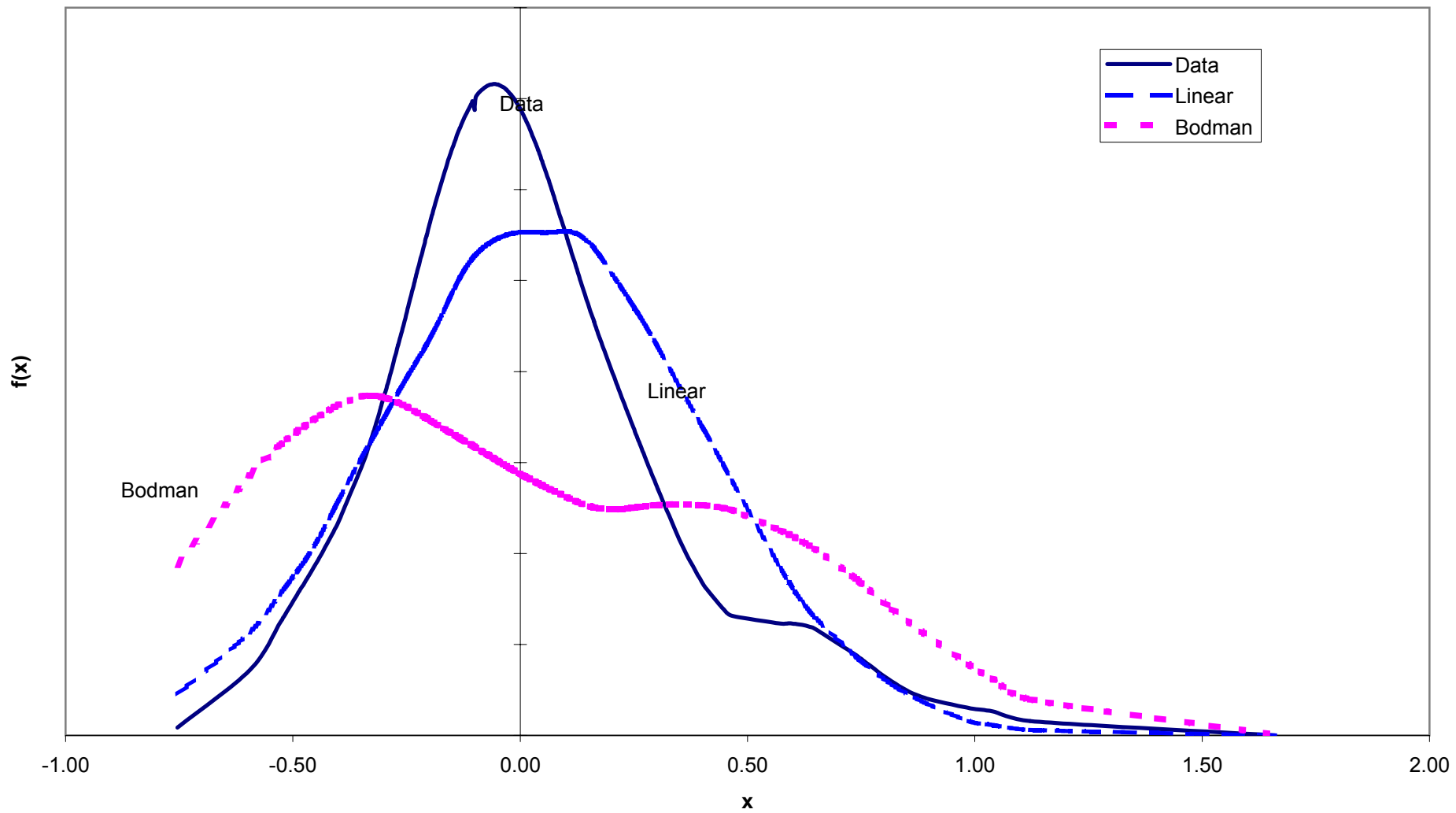


Figure 2: Density estimates for Revised MS Model

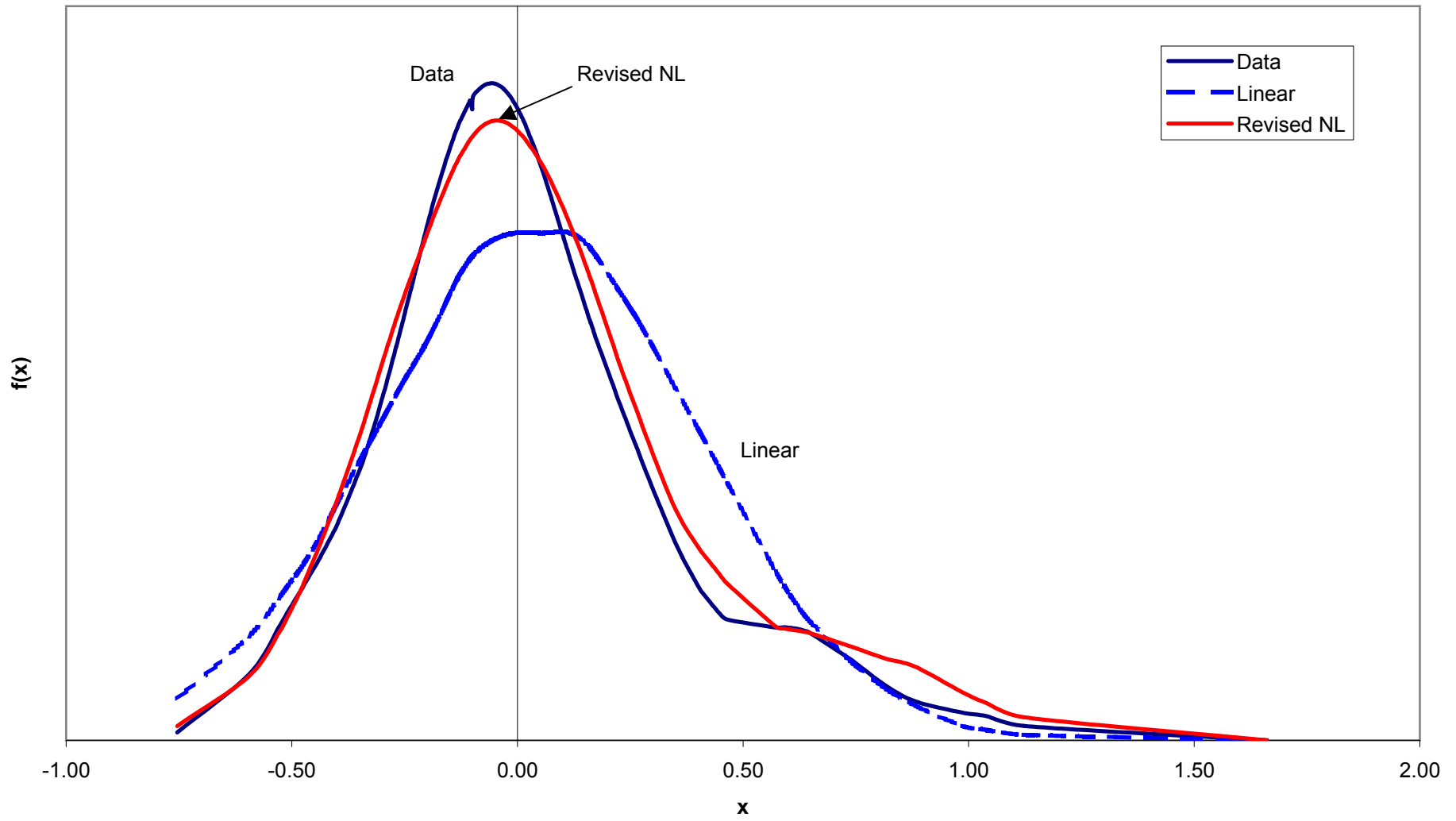


Figure 3: Change in Australian Unemployment Rate vs. Unemployment Rate Lagged

