An Index of Uncertainty for Business Cycle Leading Indicators

Minakshy Iyer
Vibhuti Vora
Rohini Khuperkar*

Abstract

Leading indicators based on correlations with reference cycles are regularly used to monitor the economy. It would be useful if we could have a quantitative measure of the risk associated with leading indicators forecasts. In this paper, we outline a methodology to develop an index for quantifying the risk of the economy actually ending up in a boom when the indicator/index predicts recession and vice-a-versa. These measures will be particularly useful for analyzing turning points, where leading indicator forecasts are at the greatest risk of going wrong. The paper carries out one exercise as an illustration and demonstrates the close correspondence between the risk function (determined in advance) and the turning points of the business cycle.

Introduction:

Leading indicators are regularly constructed in order to monitor the state of the economy and predict booms as well as recessions in advance. Typically, leading indicators are obtained by examining the cross-correlation between a reference cycle and lags of a number of candidate leading indicators. The lags of the candidate series with the highest cross-correlations (typically exceeding 0.5) are chosen, standardized for mean and variance and then aggregated into a leading indicator. This composite leading indicator index is then used to track the performance of the economy. Though a number of other methods to develop leading indicators based on coherence and mean delay and turning point analysis have been proposed, the cross-correlations methodology remains popular. [1]

---

* The authors are students of MA(II)Dept. of Economics, University of Mumbai.
The authors are grateful to Dr Neeraj Hatekar and Dr Ajit Karnik fo their valuable guidance and useful suggestions. The responsibility of remaining errors remains with the authors. Address for further correspodence: rohini_k18@yahoo.co.in

Measures have also been suggested for assessing the quality of a leading indicator, particularly its ability to capture turning points. A popular measure is given by the Quadratic Probability Score (Dieboldt and Rudenbusch (1999)). The Quadratic Probability Score (QPS) of each indicator may be used to evaluate the quality of the indicator. The QPS is defined as

$$QPS^{[H_1,H_2]} = \frac{2}{N}\sum_{t=1}^{N}(P_t - R_t)^2$$

Where P_t ($P_t = 1$ for a turning point, zero otherwise) denotes the predicted outcomes from a indicator and R_t ($R_t = 1$ for a turning point, zero otherwise) is the observed realization in the reference series. N is the total number of sample observations. By construction, the value of QPS ranges between zero and two, with zero indicating perfect prediction and two indicating no single correct signal from an indicator. [H_1, H_2] is the prediction window, which is used to determine whether a predicted outcome represents a correct signal or a false one when it takes the value of one, and whether or not it has missed a turning point when taking the value of zero. Such an indicator is useful in evaluating the quality of an indicator in an ex-post sense. However, it is not particularly useful for evaluating the efficacy of an indicator to predict turning points in advance.

In this paper, we develop an alternative methodology for assessing the quality of an indicator and measuring the risk of an incorrect prediction of the state of the economy. This methodology is outlined in section 1.

Section 1

Let us say that the economy is in state S_1 when it is in a recession and state S_2 when it is in a boom. By the risk associated with a given component of the leading indicator, we learn the risk of the economy actually turning out to be in state S_2 when the leading indicator indicates S_1 and vice-a-versa. We will measure this risk as a conditional probability, conditional on a given value of the realization of the leading indicator component.

More formally, suppose x_{t-i} is a component of the leading indicator index, with the highest correlation at lag i with the reference cycle, y_t. Suppose the cross correlation indicated here is positive. We would then expect the following:

$P(y_t = S_1 / x_{t-i}) > P(y_t = S_2 / x_{t-i})$ Whenever x_{t-i} is in state S_1

and

$P(y_t = S_1 / x_{t-i}) < P(y_t = S_2 / x_{t-i})$ Whenever x_{t-i} is in state S_2.

(Equation 1)

Since x_{t-i} is detrended with mean zero, we will say x_{t-i} is in state S_1 if $x_{t-i} \leq 0$ and in S_2 otherwise.

Given this, from Bayes' rule, we have

$$P(y_t = S_1 / x_{t-i}) = \{P(x_{t-i} / y_t = S_1) * P(y_t = S_1)\} / \{\sum_{i=1}^{2} P(x_{t-i} / y_t = S_i) * P(y_t = S_i)\}$$

And

$$P(y_t = S_2 / x_{t-i}) = \{P(x_{t-i} / y_t = S_2) * P(y_t = S_1)\} / \{\sum_{i=1}^{2} P(x_{t-i} / y_t = S_i) * P(y_t = S_i)\}$$

(Equation 2)

(Equation 1) above implies the condition

$$P(y_t = S_1 / x_{t-i}) > P(y_t = S_2 / x_{t-i}) \forall x_{t-i} \leq 0$$

And

$$P(y_t = S_1 / x_{t-i}) < P(y_t = S_2 / x_{t-i}) \forall x_{t-i} > 0$$

The risk of predicting the wrong state i-periods later on the basis of the observed realization of x_t can then be quantified as $P(y_t = S_2 / x_{t-i})$ when $x_t \leq 0$ and $P(y_t = S_1 / x_{t-i})$ when $x_t > 0$. These numbers can be read off straight away from the relevant conditional density functions. Suppose we have observed $x_t \leq 0$. Going by equation 1, S_1 is more likely relative to S_2 and hence we will predict S_1. However, the conditional probability of S_2 conditional on the observed value of $x_t \leq 0$ need not be zero. Hence, to capture this risk, we develop a measure of this risk, which is simply 1- P_1 where $P_1 =$ $P(y_t = S_1 / x_{t-i})$.

It is straightforward to generalize this to the case of n independent indicators. Suppose we have $I_i = x_i$ is the observed value for the ith indicator, $i = 1, n$. In that case, the risk of prediction of S_1 is given as follows

$$R(n) = 1 - \sum P_i / n \qquad \text{(Equation 3)}$$

If all the predictions are such that the conditional probability of S_2 conditional on the observed value of the indicator is zero for all the indicators, then $P_i = 1$ for each of the indicator and hence the measure of risk equals zero. On the other hand, if the conditional probability of S_2 given the observed value of the indicator is 1 for each of the indicator, then $P_i = 0 \forall i$ and hence, the value of risk measure is 1. Hence, we have a measure of risk that lies between 0 and 1, and can be used to evaluate the risk associated with a leading indicator forecast given various observed values of the indicator components.

In the next section, we present an example of our methodology for a simple leading indicator based on three component time series.

SECTION 2

In this section, we use the monthly index of industrial production (IIP) as the reference series. We have used monthly data on the IIP from April 1998 to March 2005 (1993-94=100). The leading indicator index comprises of aggregate deposits, term deposits and electricity production components of the index of industrial production (1993-94=100).After deseasonalising the IIP data, the partial autocorrelations and autocorrelations were examined to ensure that the data did not indicate any further seasonal fluctuation. As no significant spikes were detected at lag 12, the data were then detrended using the Hodrick–Prescott filter. The Hodrick-Prescott λ was set equal to 129600 (Ravn and Uhlig(2002)). The data were deseasonalised using a Census X-11 method. After detrending, five month centered moving average was computed to smooth the data. An identical treatment was given to components of the index of leading indicator. Cross correlations between the detrended IIP and the leading indicator index were calculated for various (1 to 24) lags of the latter and the most significant correlations were found as under

Table: 1

Leading Indicator	**Highest correlation**	**Corresponding lag number**
Term Deposits	0.82158	11
Electricity	-0.69392	19
Aggregate Deposits	0.78081	23

For simplicity, we say that IIP is in state S_1 if IIP is below its mean (which is zero) and that IIP is in state S_2 when IIP is above its mean. Ideally, we would have liked to identify S_1 and S_2 respectively with a recession and boom, but that would have resulted in too few data points for our analysis. This does not change the logic of our exercise and hence we have ignored this obvious limitation.

The risk measure that we have outlined requires us to calculate the conditional probabilities mentioned in Equation 1. Those probabilities can be calculated using Bayes' rule as in Equation 2. Suppose we want to estimate

$$P(y_t = S_1 / x_{t-i}) = \{P(x_{t-i} / y_t = S_1) * P(y_t = S_1)\} / \{\sum_{i=1}^{2} P(x_{t-i} / y_t = S_i) * P(y_t = S_i)\}.$$

We will then be required to estimate $P(x_{t-i} / y_t = S_1), P(x_{t-i} / y_t = S_2), P(S_1)$ and $P(S_2)$. $P(S_i)$ can simply be estimated as the relative frequency of the times y_t is in state S_i. The $P(x_{t-i} / y_t = S_1)$ and $P(x_{t-i} / y_t = S_2)$ can be separately estimated non-parametrically. (Chauvet and Hamilton (2006)). We explain the procedure below:

Step 1: Obtain the non-parametric density function of x_{t-i} when the economy is in state S_1, i,e, the values of x_{t-i} when y_t was below zero. The non-parametric Kernel density estimator (Rosenblatt (1956)) is given as follows:

$$\hat{f}(x_0) = (1/nh) * \sum_{i=1}^{n} K((x_i - x_0)/h)$$

Where K(.) is an appropriately chosen Kernel density estimator. (For a description of the various Kernel density estimators that can be used, see Cameron and Trivedi (2005, page.300)). We have used the Epanechnikov (or the quadratic) Kernel density estimator where:

$K(.) = (3/4)*(1-z^2)\times I(|z|<1)$ and where h is the bandwidth.

and $I(|z|<1)$ is an indicator function that takes value 1 if the absolute value of z is less than one and is zero otherwise.

The Epanechnikov kernel has an advantage over other kernels (though a slight one) in that the Epanechnikov kernel minimizes the mean integrated square error (Cameron and Trivedi (2005), page 303).

The choice of the bandwidth is critical in order to ensure the consistency of the estimates. Kernel density estimators are biased in small samples, but can be shown to be point-wise consistent at a particular point $x = x_0$ if the bias as well as the variance disappear. This is ensured if $h \to \infty$ and $nh \to 0$. In practice, one uses the Silverman's plug-in estimate given by

$h^* = 1.3643\delta N^{-0.2} \min(s, iqr/1.349)$

Where s is the sample standard deviation, IQR refers to the interquartile range and δ = 1.7188. (Cameron and Trivedi (2005), page 300).

In practice, it might be necessary to parameterize the estimated density, which can be achieved by fitting an appropriate distribution which can be determined by visual inspection of the shape of the non-parametric density (Chauvet and Hamilton (2006)). The analysis can be confirmed by use of a non-parametric goodness of fit test like the Kolmogrov-Smirnov test (Rohatgi (2003)).

Figure 1, 2 and 3 presents the estimated non-parametric density function for the three components of the leading indicator index

$P(x_{t-i} / y_t = S_1)$.

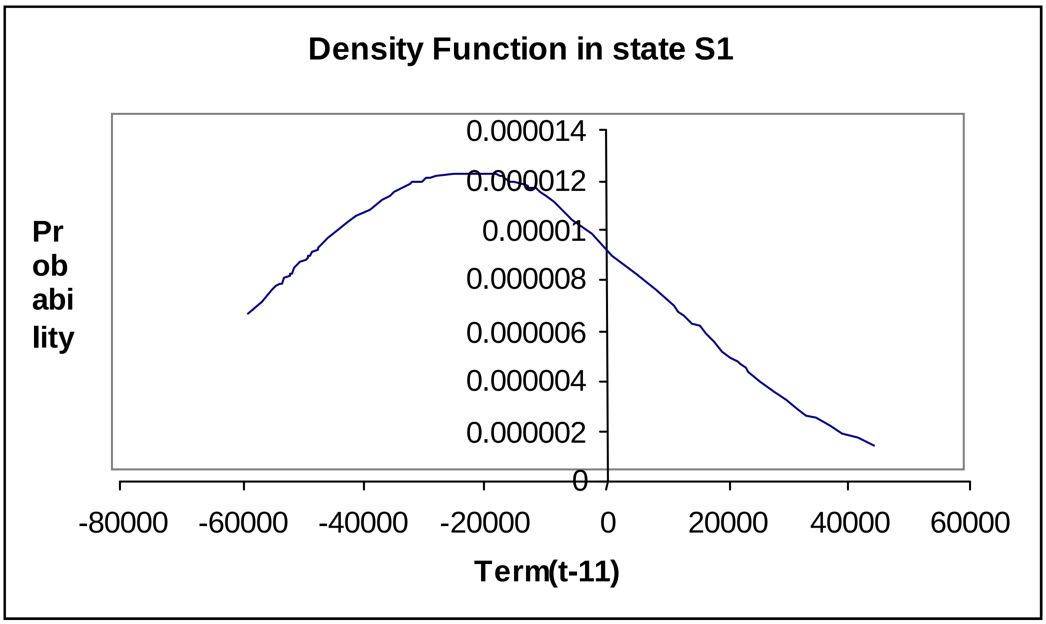
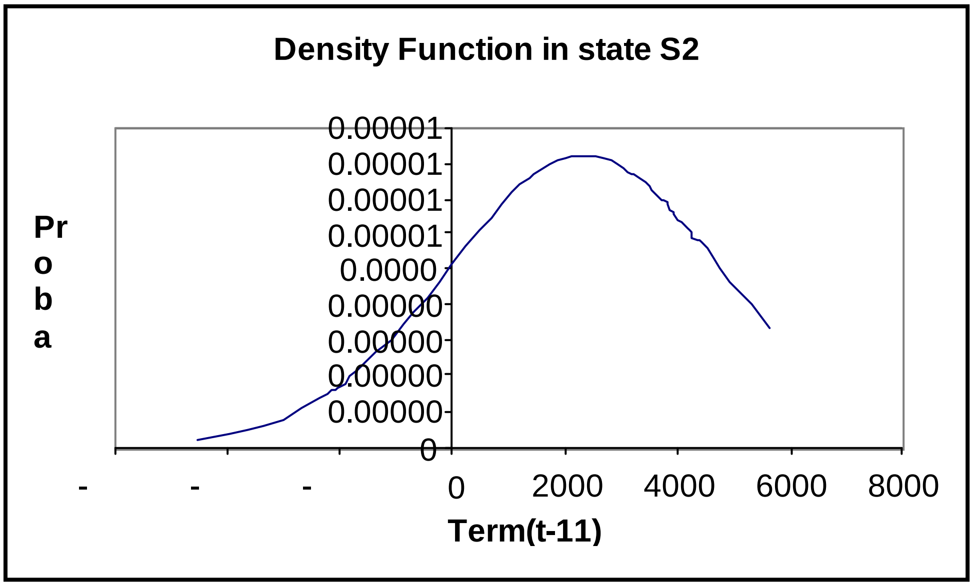


Figure 1.a (Density Function of Term Deposits in state S1)

Similarly, Figure 1.b plots $P(x_{t-i} / y_t = S_2)$ for term deposits



.Figure 1.b (Density Function of Term Deposits in state S2)

Similarly, we have state S1 and state S2 for electricity production as below:

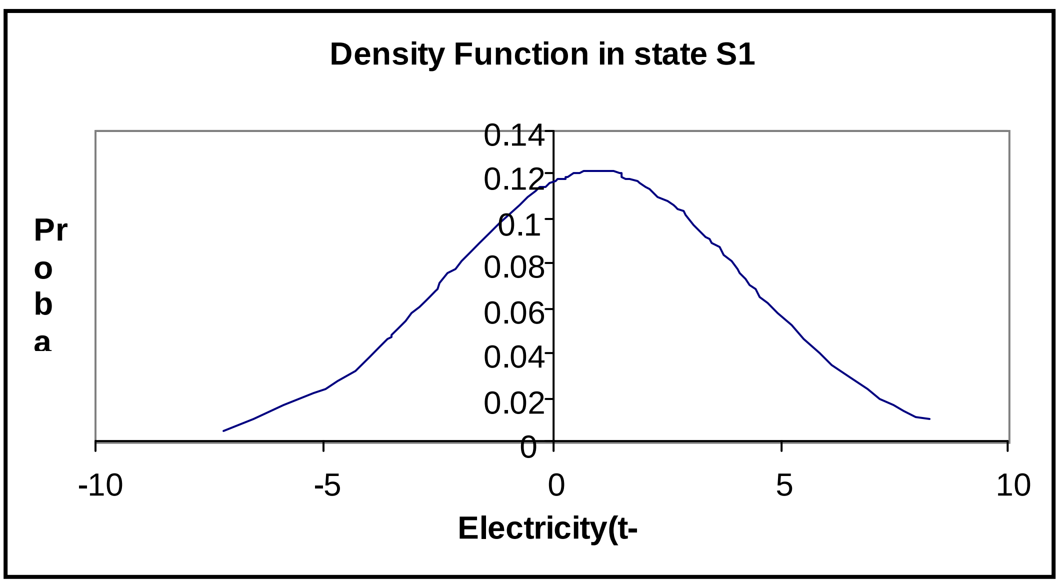


Figure 2.a (Density function of Electricity production in state S1)

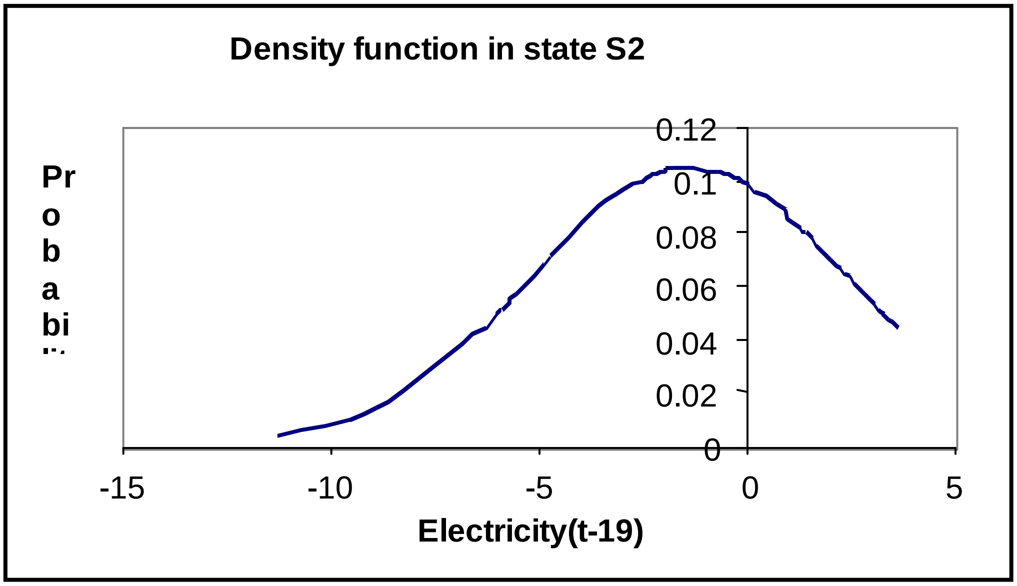


Figure 2.b (Density Function of Electricity production in state S2)

And in case of aggregate deposits we have diagrams as given below:

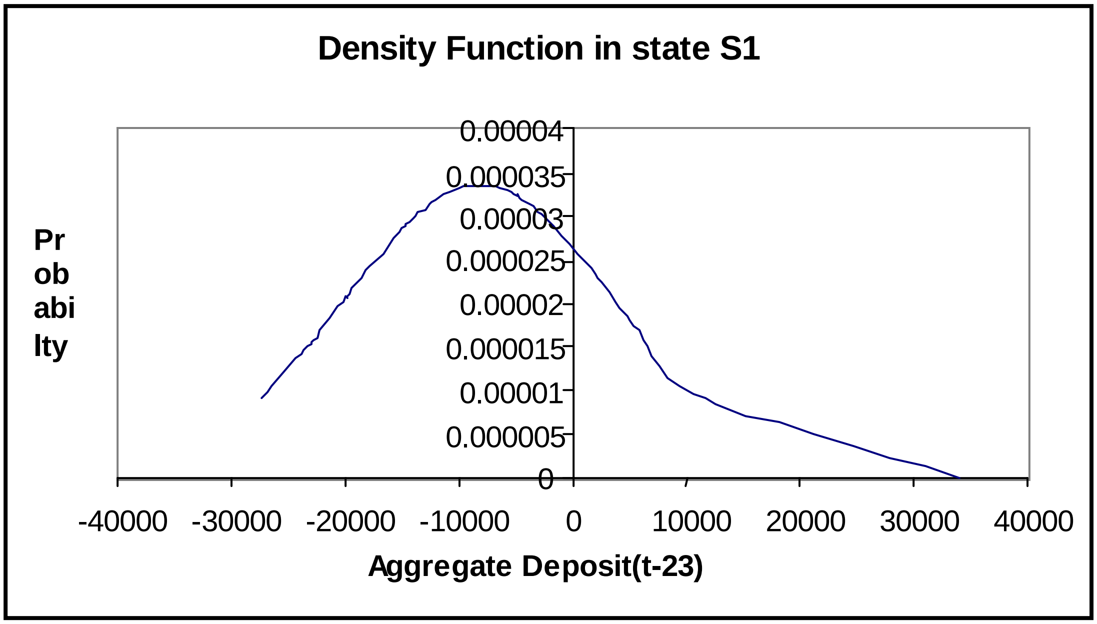


Figure 3.a (Density Function of Aggregate Deposits in state S1)

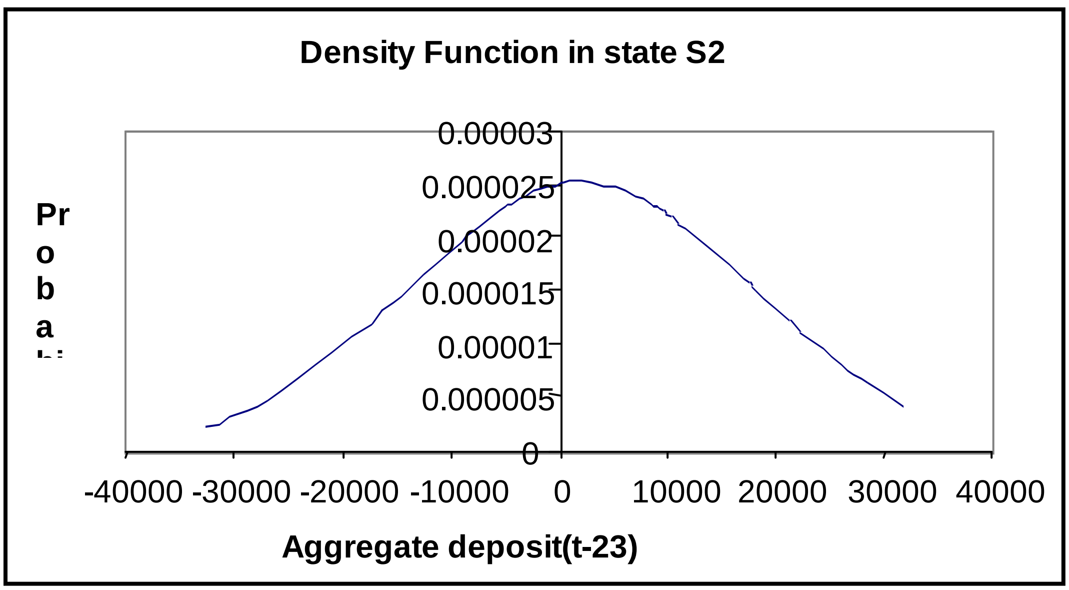


Figure 3.b (Density Function of Aggregate Deposits in state S2)

Having obtained these two density function of each component, we can then proceed to combine them using Bayes' rule, after suitable parameterization. We found that the normal density function fits the data well in each of the two states for the three indicators. The table below gives the values of the Kolmogrov-Smirnov test statistic to judge the goodness of fit for the normal density function F(x) and the observed density function $F_0(x)$ (Rohatgi(2003)).

$$H_0: \ F(x) = F_0(x)$$

$$H_1: F(x) \neq F_0(x)$$

Table: 2

Leading Indicator	***State***	***Critical value***	***Computed value***	***Decision***
Term deposits	S1 S2	0.17 0.177	0.1249 0.1459	Accept null Accept null
Electricity	S1 S2	0.172 0.193	0.0146 0.0690	Accept null Accept null
Aggregate Deposits	S1 S2	0.179 0.200	0.0282 0.0216	Accept null Accept null

Below, we combine the two density functions by applying Bayes' rule as in Equation 2 in section 1 for each of the three components of the leading indicator index.

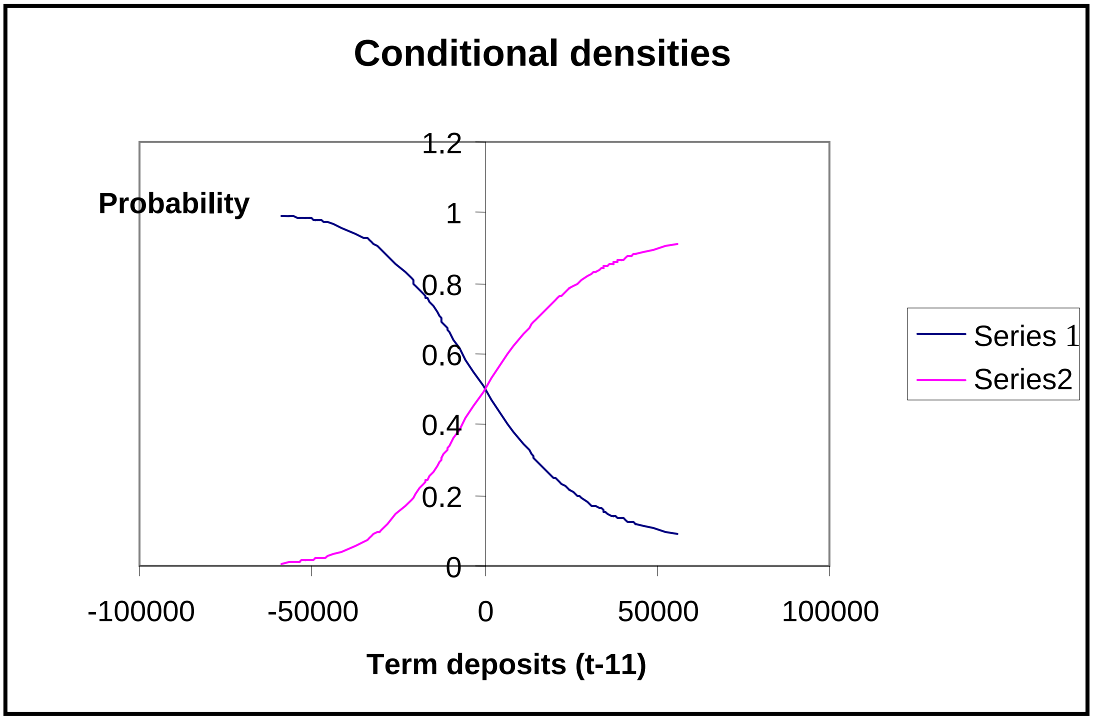


Figure 4: Term Deposits

In figure 4, the curve sloping from the left to the right gives $P(y_t = S_1 / x_{t-i})$ for various values of term deposits lagged 11 periods i.e. x_{t-11} so as to correspond with the highest correlation of IIP while the curve sloping down from the right to the left gives $P(y_t = S_2 / x_{t-i})$.

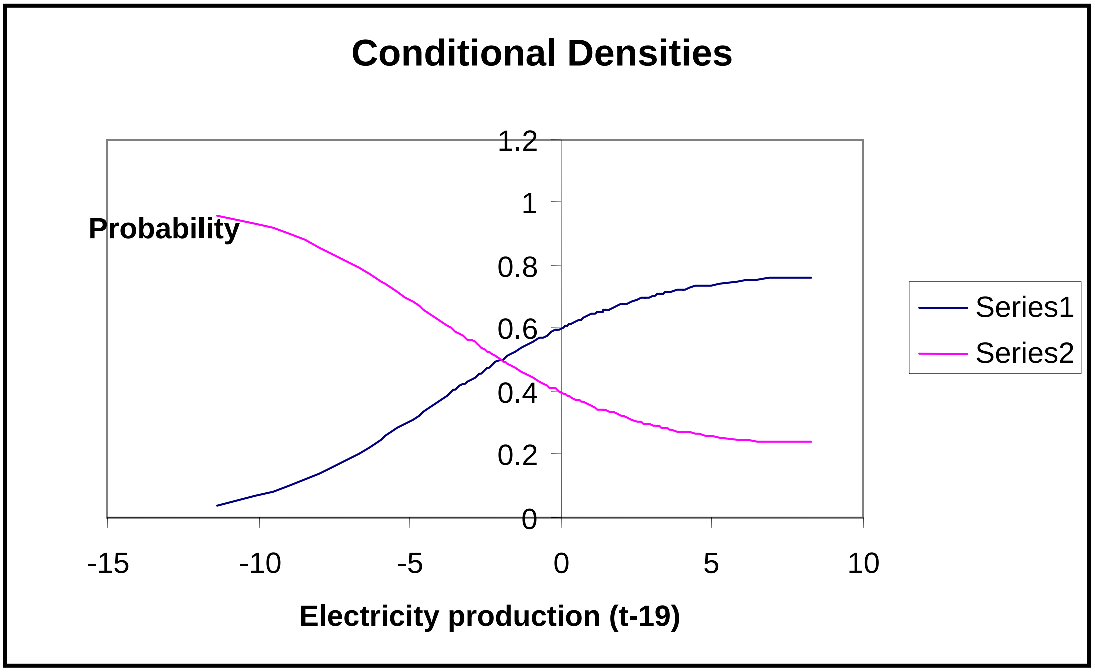


Figure 5: Electricity production

In figure 5, the curve sloping from the right to the left gives $P(y_t = S_1 / x_{t-i})$ for various values of electricity lagged 19 periods x_{t-19} while the curve sloping down from the left to the right gives $P(y_t = S_2 / x_{t-i})$. Here we must note that electricity has a high correlation of 0.69 with IIP but it is negative.

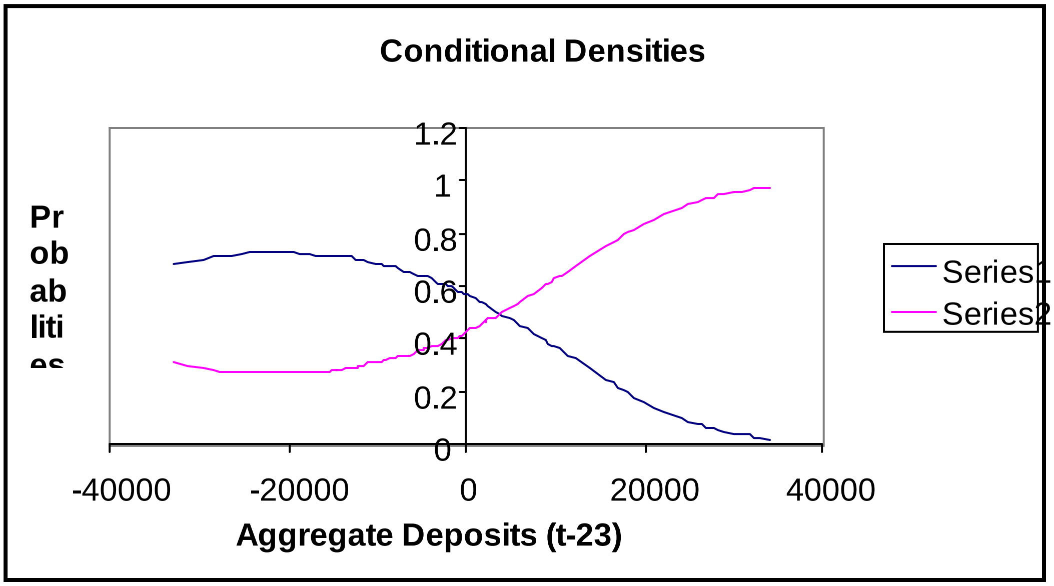


Figure 6: Aggregate deposits

In figure 6, the curve sloping from the left to the right gives $P(y_t = S_1 / x_{t-i})$ for various values of x_{t-23} while the curve sloping down from the right to the left gives $P(y_t = S_2 / x_{t-i})$.

Given these curves, suppose we have a realization of $x_{t-i} < 0$. In that case, the probability of S_1 occurring exceeds the probability of S_2 being realized at t, and hence the prediction would be S_1 . However, the probability of S_2 actually being realized at t can now be read off from the curve measuring $P(y_t = S_2 / x_{t-i})$. We must take note here that in case of electricity, since it has a negative correlation with IIP we would expect the other state to occur.

This process can easily be generalized to a leading indicator index that contains more indicators, where as has been pointed out above, risk can be measured as a function of the number of leading indicators:

R(x,n) = $1 - \sum P_i(x_i)/n$ where x_i is the observed value of the ith indicator.

Below, we plot the risk associated with prediction of S_1 using the index with components-aggregate deposits, term deposits and electricity as a leading indicator, as an illustration. We expect the risk of an actual realization of S_2 when the prediction on the basis of the conditional probabilities is S_1 to go up as one approaches a turning point when the economy shifts from S_1 to S_2. The method outlined in this chapter allows us to measure this risk and hence obtain an indication of a turning point in advance.

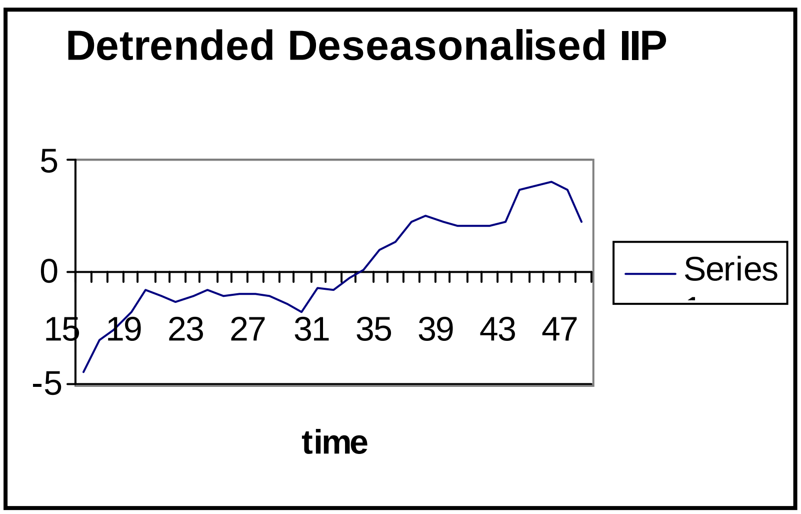


Figure 7

The above graph is that of detrended deseasonalised IIP which has been regressed on IIP with a lag of 12 periods to remove the seasonal fluctuations. The residuals thus obtained were smoothed using five period centered moving averages. This has led to a loss of 14 observations.

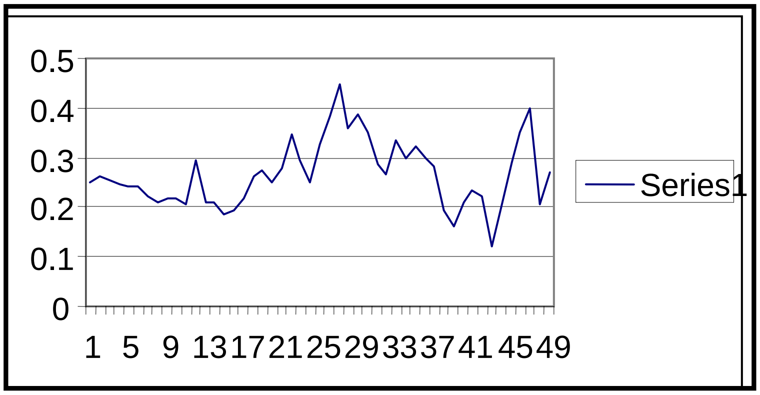


Figure 8: Risk function

The close correspondence between our risk measure and the peaks of IIP can be seen from above. The turning points in IIP can be captured by the peaks in the risk function. A number of features of the risk function are worth noting:

- From observation number 30, the index of industrial production shows a sustained rise to observation number 38. The IIP then flattens out till observation number 43, and again commences to rise till point 45. The risk function peaks at 27 and then is moving downward till it reaches point 43. In the majority of this duration, IIP has been moving upwards, indicating an expansion. During this period of stable expansion, the risk function has been moving down steadily. Though the dates do not match perfectly, this correspondence is encouraging.
- In the expansion subphase, IIP flattens out between observation numbers 38 and 44. During this phase, the risk function also goes up. After a small spurt, IIP expansion starts to slow down. During this phase, the risk function also rises.

- Between observations 19 and 21 too, IIP flattens out. Our risk function is increasing in this phase.

It is evident that the number of observations are far too few in order to reliably evaluate our risk function. However, from the limited observations that we at our disposal, the following conclusions can be tentatively drawn:

a) Whenever the economy is in a sustained expansion, the value of the risk function drops. This is to be anticipated, since the leading indicator generates a strong signal in this period.

b) Whenever the expansion is faltering , or turning around, the risk function rises.

Given that the risk function for our set of indicators can be computed atleast 11 months in advance (or more realistically, 9 months in advance given a two month lag for the data to become available), we can use this risk function in conjunction with the signals generated from the leading indicator index in order to judge the strength of the signal emitted by the leading indicator.

Conclusion

In this paper, we have outlined a statistical method based on non-parametric density estimation of quantifying the risks involved with predictions using leading indicators. Though we have chosen three components to define the leading indicator index for illustrative purposes, our method can easily be generalized to a more comprehensive index of leading indicators. This methodology can also be used then to indicate possibility of turning points.

BIBLIOGRAPHY

Cameron A.C and P.K.Trivedi (2005): Microeconometrics: Methods and Applications, Cambridge University Press, Cambridge.

Chauvet M. and J.D. Hamilton (2006): "Dating Business cycle Turning Points" in Milas C., Rothman P and D. van Dijk (eds). Nonlinear Time Series Analysis of Business Cycles, Elsevier, Amsterdam, pp.1-54.

Diebold, F.X. and G.D. Rudenbusch (1999): Business Cycle – Durations, Dynamics and Forecasting, Princeton University Press, Princeton, N.J.

Ravn M.O and H. Uhlig(2002) : "On Adjusting the Hodrick-Prescott Filters for the Frequency of Observations", The Review of Economics and Statistics, vol. LXXXIV, May 2002, No.2, pp.371..376.

Rohatgi. V. (2003) : Statistical Inference, Dover publications Inc, New York, pp. 754..760

Rosenblatt M. (1956): "Remarks on Some Non-parametric Estimates of a Density Function", Annals of Mathematical Statistics, 27, pp.832-837