Knowing A Bit but Not Too Much: Incomplete Directional Models and Their Use in Forecasting and Hedging

Roger Bowden, Jennifer Zhu and Jin Seo Cho

Victoria University of Wellington
School of Economics and Finance
PO BOX 600, Wellington, NZ

January 2006

Abstract

Directional calls are often more successful than precise value prediction, particularly at certain times, when underlying fundamentals suggest a breakout from the stable range. We adapt the categorical directional framework implicit in binomial or trinomial step processes to establish nonhomogeneous multinomial directional probabilities over coarser time intervals and show how such frameworks can be used for forecasting and hedging, including dynamic persistence. Problems of signal compression and outcome definition can be addressed using methods analogous to neuronal nets and fuzzy membership functions. The methods are applied to derive forecasting and conditional hedge procedures for foreign exchange exposures.

JEL classifications: C22,C25,C32,C51,C53; E27; F31, F37; G13; M21.

Key words: **Conditional value at risk**, **foreign exchange forecasting**, **fuzzy regimes**, **hidden Markov models**, **non-homogenous multinomial process**.

* Thanks go to David Harte for some useful references also to Halbert White and other participants in the Second Workshop on Hidden Markov Models and Complex Systems held in December 2005 at Victoria University of Wellington. Emails: Roger.Bowden@vuw.ac.nz; Jennifer.Zhu@vuw.ac.nz; JinSeo.Cho@vuw.ac.nz.

1. INTRODUCTION

Directional forecasting is concerned with calls as to whether a given series will move up, stay the same, or move down, preferably by means of assigning numerical probabilities to the outcomes, given the available information. Assigning directions rather than actual values is a natural response to limited information situations where managers do not have any very precise structural or casual model, as is commonly the case. Thus business cycle commentators are willing to call only the basic direction of movement, based on implicit theorizing as to choice of indicators, and abandoning any pretence that they are able to answer very precisely the 'when' and 'by how much' questions. Consumer or producer confidence surveys likewise ask respondents for directional assessments, and how strongly they feel about them - they do not ask respondents to guess an actual value. Central bankers tailor their monetary policy announcement in similar terms, for example signalling that future cash rate tightening may be necessary, without saying just how much. Even in supposedly efficient financial markets, there is some evidence that directional forecasting may be successful (e.g. Leung, Daouk and Chen 2000; Levich 2001; Christofferson and Diebold 2004; Pesaran and Timmermann 2004), no matter that value forecasting is not.

The categorical directional framework is a much used dynamic modelling device in a number of other contexts. In security pricing, binomial or trinomial jump processes are used to approximate Brownian motion or other continuous time processes, where the time interval is small. In the case of nonhomogeneous jump processes, the jump probabilities can be made to depend upon observable state variables or exogenous variables, and the methods of the present paper can certainly apply to modelling of this kind. In most forecasting or risk management situations, however, data limitations imply relatively coarse time intervals: a month, quarter or year, so that one is no longer necessarily thinking in terms of continuous time approximations. The directional models that result may be viewed as nonhomogeneous multinomial processes.

Managers might also feel that if a directional probability is high, the value of the movement itself is likely to be appreciable and should be acted upon. It does not automatically follow that this should be the case – e.g. the probability could be higher simply because the variance is lower, without materially affecting the expected size of the movement. Nonetheless, if the probability of a rise in the home exchange rate is assessed as 0.9 say, then export managers would load up on foreign exchange forwards to protect their foreign currency receipts, on the grounds that the size of the movement is also likely to be significant. The general problem is how to map directional calls and probabilities into actions. A first step is to establish a suitable welfare or loss function associated with possible outcomes. Loadings can then be devised that weight the directional probabilities according to the welfare consequences of state transitions, so that the resulting hedge ratios reflect not only directional probabilities, but also welfare outcomes. The resulting decision problem can also be used to inform the statistical inference associated with variable selection in the estimation phase. The link between forecast and action appears in the literature on the choice of loss function for forecast evaluation, including limitations of the mean square error criterion (e.g. Leitch and Tanner 1991; Diebold and Mariano 1995; Leung *et al* 2000). In a directional context, it is necessary to go beyond this to devise decision rules that map directional probabilities into actions.

As earlier indicated, directional forecasting is often a response to limited but still useful information. One problem about partial information is that it may be technically incomplete. The US Conference Board (TCB) business cycle indicators provide an illustration, with their broad classification into leading, coincident and lagging indicators, each made up of a number of economic variables. The leading indicator is used to describe when the economy looks like coming out of a recession, and the lagging indicator when the economy looks like slowing down. This is not to say that the leading indicator is altogether useless in detecting a slowdown, just that the focus of attention has shifted. So these are partial indicators: one could think of the leading series as the ‘up state’ signal and the lagging series as the ‘down state’ signal. They can also conflict, so that an ‘up’ signal could potentially light up at the same time as a ‘down.’ Technically this would mean that for some configurations of the indicator signals, there could be multiple possible outcomes, and the observer has to choose which seems the most probable based on judgements such as the relative strengths of the signals, or by some simple randomization device. The lack of a unique functional relationship mapping between the signal space and the outcome space is what makes the information incomplete. The word ‘incomplete’ can also be used with respect to missing data in the estimation phase (see below).

The purpose of the present paper is to establish procedures for directional forecasting with incomplete information, and to show how to adapt the forecasts to an underlying welfare function. The standard model has two underlying signals, which we can characterize as respectively up and down in orientation, and either two outcomes (up/down) or three outcomes (up/no change or stable/down). The signals are noisy signals, and this being the case it is quite possible that they conflict, so that in general there are 4 possible signal configurations, which have to be mapped into either the 3 or 2 observable outcomes. Decision theorists would say that the signals lack 100% validity. In addition, the signals are themselves not directly observable. Instead one has available a number of observable economic variables that one can use as signal indicator variables. In present terminology, the TCB leading, coincident and lagging business cycle indicators would in fact be signals, and the economic variables that go toward making them (M2, inventories, employment etc.) would be the indicator variables. One can say that these are noisy signals, noisily observed. There are similar frameworks in the economics of incomplete information, but the primary objective in the present paper is conformance with the common predicaments of commercial life. To be sure, there may be other agendas to the use of such models. For instance, they are highly nonlinear and can be used to re-examine issues of capital market efficiency; assuming, of course, that the market does not already know about them.

The resulting model can be regarded as a sequence of multinomial trials (3 outcomes) or binomial (2 outcomes), one trial for each time period. At each time period the probabilities change in accordance with economic conditions, so that these are nonhomogeneous multinomial models. The probabilities themselves have to be derived by mapping the indicators into probability densities. This can be done by using standard distribution models such as the logistic or normal probit, supplemented with rule-based judgment calls where the signals can conflict. There are useful commonalities with the neuronal (neural) network literature, which involves a similar compression process (e.g. McCulloch and Pitts 1943; Rosenblatt 1957, 1958; Kuan and White 1994; Turban 1995). Most of our models can be cast as neuronal nets, with several layers. The three-outcomes become collectively a hidden feed forward layer if the primary objective is a two outcome model, so the parallel with neuronal nets becomes even closer. However, a greater degree of prior structure is imposed than would be normal with the neuronal net models, with the purpose of contributing data economy within a macroeconomic context. Most economic or financial applications of neuronal nets are concerned with shorter run forecasting with high frequency data, though for an exception see Tkacz and Hu (1999) on GDP forecasting.

A further contribution arises in connection with estimation of categorical models in the presence of hidden data, where we note and exploit a connection between the EM algorithm, on the one hand, and the literature on fuzzy membership functions (Zadeh

1965), on the other. Our results suggest that it is much more effective to have three outcomes, allowing for a middle ground of ‘neither up nor down’, as the old song goes. Absent precisely zero outcomes (as with most continuous variates), the operational problem is then how to define the ‘no change’ or ‘stable’ outcome. The boundaries separating categories can be regarded as fuzzy, and the parameters of the membership function estimated from the historical data. A probabilistic interpretation can be given in terms of the regime boundaries as hidden variables. One replaces the unobservable regime membership functions with their expected value, given the available information, in a fashion similar to the EM algorithm (Hartley 1958; Dempster, Laird and Rubin 1977). This establishes an interesting link between the use of fuzzy regimes and likelihood methodology with incomplete data, potentially useful in other forms of categorical data analysis.

Finally, the model can be either static or dynamic, though in both cases the probabilities change over time. In the latter case, directional changes can themselves be correlated over time, so this is a model of dynamic directional persistence. The important probabilities are transitional and the resulting model can be interpreted as a variant of the hidden Markov model (Baum and Petrie 1966; Bickel, Ritov and Ryden 1998; Cappe, Moulines and Ryden 2005).

The scheme of the paper is as follows. Section 2 is concerned with underlying probability structure. Incompleteness is identified with indicator conflict and ways to resolve this are described. The model is extended to dynamic persistence via the use of nonhomogeneous Markov processes. Section 3 moves on to discusss operational matters. In the case of three outcomes from continuous data, the no-change or stable regime has to be defined or delineated, and this issue is addressed with the use of ideas from fuzzy logic, which can be given a probabilistic interpretation in terms of hidden data. Estimation and testing procedures are established, in the form of quasi-maximum likelihood. Section 4 contains the empirical application. It is shown in Section 4.1 that variables such as house prices, relative share prices, or the current account balance are collectively superior to the forward rate in calling directional changes in the NZ dollar v. US dollar exchange rate. The application to hedging is developed in Section 4.2. A methodology is developed to embody directional probabilities in hedge ratios. The optimized hedge ratios represents an improvement on both the unhedged exposure and the simple forward. Section 5 has some concluding remarks.

2. THEORY

The initial objective in directional forecasting and risk management is to attach probabilities to categorical outcomes. The outcome space can be either two or three dimensional. In the former, the categories are simply ‘up’ or ‘down.’ In practice, however, it is useful to have a middle category which, depending on context, can be described as ‘stable,’ ‘no change,’ or ‘same.’ Given that most economic or financial times series exhibit noise or normal volatility, one would be reluctant to accept smaller movements as a true up or down trend. Technical analysis of markets uses a similar idea in the form of a ‘break out’ zone. Both two and three dimensional outcome spaces are considered in what follows. The problem of how to define the output zones in practice is considered in the next section.

The forecasting is to be based on a series of economic signals. Attention is usually focussed on signals that might indicate a significant up or down movement. There are several reasons for this. One is that busy managers are more likely to focus on ‘heads up’ information, i.e. real news. Another is that economic models, mental or otherwise, tend to be more convincing in describing significant changes in state. Thus most commentators would accept that an unexpected announcement of a record current account deficit is likely to lead to a down movement in the home exchange rate, or that in a small economy a bullish housing market is likely to lead to an up movement as capital is hoovered in from abroad to fund mortgages. In general, the informational content of economic signals is greatest in describing up or down movements. This suggests that the natural way to proceed is in terms of a two dimensional signal space and a two or three dimensional output space. The signals can be taken as sufficient statistics for a larger number of more elementary economic indicator variables, in a manner to be described. However, the signals themselves can be observed only with noise, so they are latent variables.

It is helpful to begin by imagining that there are two signals I_1 and I_2, each observable, reserving for later discussion the noisy observation aspect. We shall call one the up signal and the other the down signal. However these are only indicative signals. There is signalling noise involved, or signal credibility, so that it is quite possible for an up signal to be followed by a down outcome. Of course for the signals to have validity we should not expect this to be a normal state of affairs, but we should allow for the possibility, simply because economic signs can be misleading.

The signals each have two states, namely 'on' and 'off.' In figure 1 these are marked with ( ) and (0) signs. Consider first the three outcome model. With two signals there are four possible combinations. It seems clear enough that if the up signal is on and the down signal is off, the outcome should be an up state, and vice versa. If both signals are off, then one would be confident in assigning the 'no change' outcome. Suppose however, that both the up and down signals are on. There is no automatic or obvious way to resolve the conflict (marked with crossed cavalry swords). In principle, the outcome could be any one of the three possible outcomes R_i as marked, though this is not to say that the strength of the linkages need to be the same. In this sense, the model is incomplete. Note that figure 1 is of the nature of a logic diagram. It could be reworked as a neuronal net: the observable economic variables (**Z**'s) feed forward to the *I*-combinations, with ongoing links (some of zero strength) to the *R*'s, and thence to the outcomes.

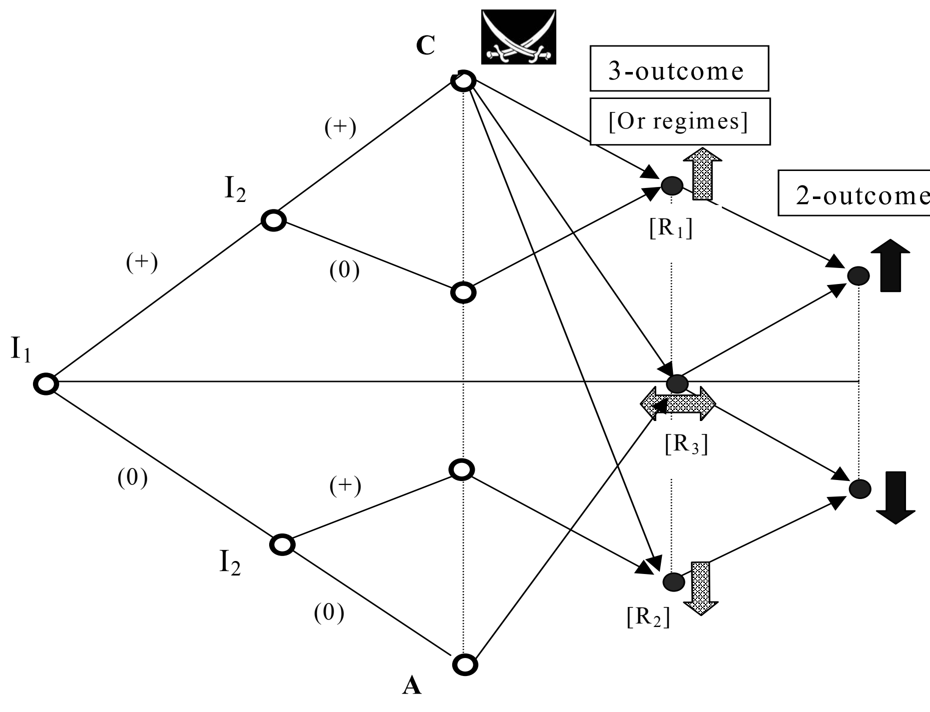

Figure 1: **Three and Two Outcome Incomplete Models**

Consider next the two outcome model. In this case, one does not observe the no change state. For instance, with continuous observations one might argue that 'no change' represents an outcome of precisely zero change, and therefore has zero probability. However, it is useful to keep the three original outcomes but to squash them down to two. The original three we could now refer to as regimes, and in the absence of direct observation, they are latent. Denote them as R_1 (up), R_2 (down) and R_3 (no change). Collectively they form a hidden feedforward layer, in the language of neuronal nets. The incompleteness arises from the fact that given R_3, there is no automatic way to assign the final up or down outcome.

Figures 2 (a,b) show by way of contrast how complete models would look in each case. The two-outcome model would be complete only if the two signals were mutually exclusive, so that an up state in one automatically means a down state in the other.

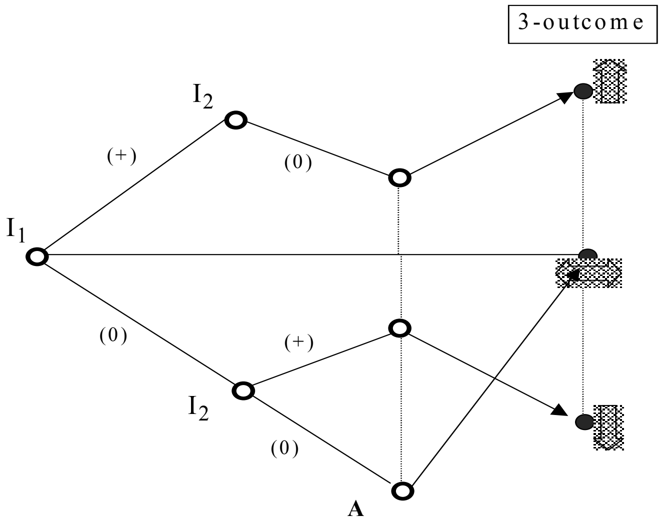

Figure 2: **(a) Complete Three Outcome Model**

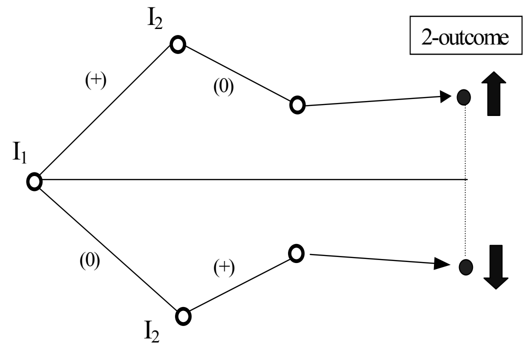

Figure 2: **(b) Complete Two Outcome Model**

Whether it is best to specify a two or three outcome model is a matter of context and purpose. Economic variables, especially financial ones, are intrinsically noisy or volatile, so that one might be unwilling to identify smaller movements with up or down changes. Instead there would be a range such that if the movement exceeded this, it would be labelled as either an up or a down change. One could think of it in terms of accepting or rejecting the null hypothesis of no change and the regime boundaries as the critical points. This would indicate a three regime model, and either using the data to determine the critical points or else defining the regimes in terms of fuzzy membership functions. Operational procedures are considered further below.

2.1 Introducing Probabilities

Allocation rules for incomplete models can be devised in terms of probabilities, which can be used more comprehensively to describe the *strengths* of connections in the above diagrams. In this respect, the basic task is to determine the marginal or conditional probabilities attached to the three regimes. We will take it that even for the two outcome model, the hidden layer regime membership variables R_i ($i \quad 1,2,3$) are sufficient statistics, so that for any observable exogenous indicator variables $\mathbf{Z}$, and outcomes u or d, we must have $P\ u|R_i,\mathbf{Z} \qquad P\ u|R_i\ P\ R_i|\mathbf{Z}$, similarly for d. Hence the essential task in either model is to make probability statements about the regime membership variables R_i.

To assist in this task it is assumed that two sets of observable economic variables are available, denoted $\mathbf{Z}_u$ and $\mathbf{Z}_d$, sometimes collectively as $\mathbf{Z}$. The two sets are oriented respectively towards up and down outcomes, e.g. a higher value of a $\mathbf{Z}_d$ variable would suggest a higher likelihood of a down movement. They can have elements in common, provided that model consistency or identification are preserved; for example a given economic series could appear both as an up indicator variable and with a negative sign, as a down indicator variable. Where directional movements are concerned, it is common to focus on variates that are adapted towards significant movement, rather than those that might specifically indicate a stable or no change outcome. The latter is more naturally thought of as associated with the absence of indications as to strong upward or downwards pressures. Hence one problem is to explain how just two sets of observable indicator variables can explain three outcomes in a natural manner. Note also that the orientation of any indicator variable is by no means perfect; wrong indicator signals can be given on occasion. For many purposes is convenient to think in terms of linear combinations $\ _u'\mathbf{Z}_u$ and $\ _d'\mathbf{Z}_d$ of these variables as producing a closer link to the respective regimes or observable outcomes, but again, wrong signals can be given.

There are just two stochastic index functions, denoted I_1 and I_2 for the indicator signals. These will depend upon their respective indicator variables $\mathbf{Z}_u$ and $\mathbf{Z}_d$ but we will often suppress these for simplicity. Each has two alternative symbolic values I_i and I_i^0; the indicating that the signal is switched on, and the 0 indicating that it is switched off. The first task is to determine the probabilities $P\ R_i|\mathbf{Z}$ to be attached to the three regimes.

Probit-style switching models provide a useful illustration. Let $\ _u$ and $\ _d$ be two random variables with mean zero and unit variance, and further let $W_1\ \mathbf{Z}$ and $W_2\ \mathbf{Z}$ be $P\ I_1 \quad I_1|\mathbf{Z}$ and $P\ I_1 \quad I_2|\mathbf{Z}$ respectively. If we assumed that $\ _u$ and $\ _d$ were independently normal, we might specify

$$\begin{aligned} W_1\ \mathbf{Z} &\quad P\ \ _u \leq \ _u'\mathbf{Z}_u|\mathbf{Z} \qquad \ _u'\mathbf{Z}_u\ ; \\ W_2\ \mathbf{Z} &\quad P\ \ _d \leq \ _d'\mathbf{Z}_d|\mathbf{Z} \qquad \ _d'\mathbf{Z}_d\ , \end{aligned} \tag{1}$$

where denotes the standard normal distribution function. Hereafter, we abbreviate $W_1\ \mathbf{Z}$ and $W_2\ \mathbf{Z}$ as W_1 and W_1 for simplicity.

Probabilities such as W_1 and W_2 will sometimes be denoted the 'raw scores' in what follows. Alternatively we might have chosen the logistic distribution, much used in studies of modal choice. A user who prefers this option could let W_1 be

$$\frac{1}{1 \quad \exp - \ _u'\mathbf{Z}_u}. \tag{2}$$

Similarly for W_2. Or we could relax the assumption that $\ _u$ and $\ _d$ are uncorrelated. This is useful in testing whether regime decompositions do in fact apply (see below).

Assuming for expositional purposes the normal model (1), there are four possible combinations of signals:

I_1, I_2^0 unequivocally indicating up, with probability $W_1 \; 1 - W_2$;

I_1^0, I_2 unequivocally indicating down, with probability $W_2 \; 1 - W_1$;

I_1, I_2 conflict zone both indicating up and down, with probability $W_1 W_2$;

$I_1^0, \; I_2^0$ agree on 'no change' outcome, with probability $1 - W_1 \quad 1 - W_2$.

(3)

More generally, if $_u$ and $_d$ are correlated,

$$P \quad I_1, I_2^0 \, |\mathbf{Z} \qquad \int_{{}'_d\mathbf{Z}_d}^{{}'_u\mathbf{Z}_u} n \quad {}_u, \; {}_d \; d \; {}_u d \; {}_d,$$

where $n \quad {}_u, \; {}_d$ denotes the joint density with the required correlation, say. Similar expressions hold for the other three combinations.

The four probabilities associated with combinations have to be compressed to the three regime probabilities $P \; R_i|\mathbf{Z}$, $i \quad 1,2,3$. It will be apparent from figure 1 that, for instance, $P \; R_1|\mathbf{Z}$ is at least $W_1 \; 1 - W_2$, but it could well be more, as mass from point C still has to be distributed. Thus we could write

$$P \; R_1|\mathbf{Z} \qquad W_1 \; 1 - W_2 \qquad f_1 \; W_1, W_2 \; W_1 W_2,$$

where $f_1 \; W_1, W_2$ is a value between zero and unity indicating the fraction of the unresolved probability mass $W_1 W_2$ to be redirected towards regime 1. There are corresponding statements for $p \; R_2|\mathbf{Z}$ and $p \; R_3|\mathbf{Z}$. The redistributive semipositive fractions f_i have to add up to unity and should be symmetric, i.e. $f_2 \; W_1, W_2 \qquad f_1 \; W_2, W_1$.

A conditional probability framework is a useful way of generating redistribution fractions with the required properties. Suppose the third combination in (3) holds, i.e. conflict. We could interpret $f_i \; W_1, W_2$ as the conditional probability $P \; R_i | \; I_1, I_2 \; , \mathbf{Z}$ of R_i, given signal conflict. A convenient specification is

$$\begin{aligned} f_1 \; W_1, W_2 \qquad P \; R_1 | \; I_1, I_2 \; , \mathbf{Z} \qquad & W_1 / \; 1 \quad W_2 \; ; \\ f_2 \; W_1, W_2 \qquad P \; R_2 | \; I_1, I_2 \; , \mathbf{Z} \qquad & W_2 / \; 1 \quad W_1 \; ; \\ f_3 \; W_1, W_2 \qquad P \; R_3 | \; I_1, I_2 \; , \mathbf{Z} \qquad & 1 - \frac{W_1}{1 \;\; W_2} - \frac{W_2}{1 \;\; W_1} . \end{aligned} \tag{4}$$

It is easy to show that $0 \leq f_i \; W_1, W_2 \; \leq 1$ and $\sum_i f_i \quad 1$. If the raw score W_1 for the up indicators is greater than that for the down indicators, it is more likely that any conflict would be resolved in favor of up. If both W_1 and W_2 tend to unity, the first two conditional probabilities (3) tend to $\frac{1}{2}$, indicating that one or other of the strong opinions must be correct, and it is just a matter of tossing a coin. If W_1 and W_2 are both $\frac{1}{2}$, then the three conditional probabilities have value 1/3 each, so the mass in the conflict zone is spread equally.

Expressions (4) are not the only possible way to divide the conflict mass; for instance, we could have chosen $P \; R_1 | \; I_1, I_2 \; , \mathbf{Z} \qquad \frac{W_1}{1 \;\; 2 W_1 W_2}$ and still achieved conditional probabilities all lying between zero and unity. The latter would say that when W_1 and W_2 are both equal to unity, then the mass is distributed 1/3 each across the three regimes.

The conditional probability framework captures the common sense of credibility judgements: if signals conflict, one or both of them must be wrong and one has to weigh up which, in the light of all available evidence. However, it does introduce a non Markovian element into the logic, for the indicator variables **Z** can now reach forward to influence signal credibility as well as signal probabilities.

Combining (4) and (3) we get

$$\begin{aligned} &P\ R_1|\mathbf{Z} \quad\ W_1\ 1 \quad W_2\ W_1 - W_2\ \ /\ 1 \quad W_2\ ; \\ &P\ R_2|\mathbf{Z} \quad\ W_2\ 1 \quad W_1\ W_2 - W_1\ \ /\ 1 \quad W_1\ ; \\ &\quad P\ R_3|\mathbf{Z} \quad\ 1 - P\ R_1|\mathbf{Z}\ - P\ R_2|\mathbf{Z}\ . \end{aligned} \tag{5}$$

Squashers like (5) obey the elementary symmetry axiom that if the raw scores W_1 and W_2 are equal, then the up and down probabilities are likewise equal. As earlier noted, this is by no means the only way of generating the multinomial probabilities required. However it is a simple way to do so, and can readily be adapted for other distributions - thus the logistic specification could replace the probit specification in the above.

It may be remarked that the role of the economic indicator variables $\mathbf{Z}_u$ and $\mathbf{Z}_d$ is to assist in discriminating between up/down/stable directions. One might think of extending their informational role into values as well as directions, in a manner parallel to Tobit models, with a statement like

$$E\ Y|\mathbf{Z} \qquad {}_u^{\prime}\mathbf{Z}_u\ P\ R_1|\mathbf{Z} \qquad {}_d^{\prime}\mathbf{Z}_d\ P\ R_2|\mathbf{Z}$$

assigning the value zero to the expected value given R_3. There are two problems with this. First, the coverage of $\mathbf{Z}_u$ and $\mathbf{Z}_d$ may not extend as far as forecasting actual values - they are simply direction indicators. Second, the probability structure is incomplete, so that the above expectation expression lacks a sound foundation in probability theory. Such expressions will not be used in what follows.

2.2 Two Outcomes

A further squashing to a two outcome model, if desired, can again use conditional probabilities. We divide the 'stable' outcome half each to the up and down outcomes: $P\ u|R_3 \quad P\ d|R_3 \quad \frac{1}{2}$. Also set $P\ u|R_1 \quad 1$, and $P\ d|R_1 \quad 0$, similarly for the R_2 conditionals. Then

$$\begin{aligned} &P\ u|\mathbf{Z} \quad \tfrac{1}{2} \quad \tfrac{1}{2}\ P\ R_1|\mathbf{Z}\ - P\ R_2|\mathbf{Z}\ \ ; \\ &P\ d|\mathbf{Z} \quad \tfrac{1}{2} \quad \tfrac{1}{2}\ P\ R_2|\mathbf{Z}\ - P\ R_1|\mathbf{Z}\ \ . \end{aligned} \tag{6}$$

For the model based on (3) we have $P\ u|\mathbf{Z} \quad \frac{1}{2}\ 1 \quad W_1 - W_2\ \ 1 \quad W_1 W_2$ to a fairly good approximation.

One could allow for the mapping between regimes and outcomes to be less precise, for example by writing

$$\begin{aligned} &P\ u|R_1 \qquad ;\ P\ u|R_2 \quad 1 - \ \ ;\ P\ u|R_3 \quad \frac{1}{2}; \\ &P\ d|R_1 \quad 1 - \ \ ;\ P\ d|R_2 \qquad ;\ P\ d|R_3 \quad \frac{1}{2}. \end{aligned}$$

The first expression in (6) is replaced by $P\ u|\mathbf{Z} \quad \frac{1}{2} \qquad - \frac{1}{2}\ \ P\ R_1|\mathbf{Z}\ - P\ R_2|\mathbf{Z}$, with a similar second expression. The effect of $\quad 1$ is to allow the connections between regimes and outputs to be noisy. Thus a validity ratio for R_1 as a signal of u versus d would be

$$v_1 \quad \frac{P\ R_1\,|\,u,\mathbf{Z}}{P\ R_1\,|\,d,\mathbf{Z}}.$$

The validity ratio for R_1 is high if a subsequent up outcome is much more strongly associated with a prior occurrence of R_1 than is a subsequent down outcome. Using Bayes' theorem,

$$v_1 \quad \frac{P\ u|R_1,\mathbf{Z}}{P\ d|R_1,\mathbf{Z}} \quad \frac{P\ d|\mathbf{Z}}{P\ u|\mathbf{Z}},$$

with a similar validity ratio v_2 for R_2 as a signal for d. Now if $P\ R_1\,|\,\mathbf{Z} \quad P\ R_2\,|\,\mathbf{Z}$, then $P\ d|\mathbf{Z} \quad P\ u|\mathbf{Z}$ and

$$v_1 \quad \frac{}{1-} \quad v_2.$$

The validity or signal to noise ratio is highest if 1. One would prefer $\frac{1}{2}$. Simulations suggest that allowing 1 does diminish the sharpness of model identification, and we suggest setting 1 unless there are strong reasons to suppose otherwise.

Some empirics reported below revealed that the three outcome model is much superior to the two outcome model in the case of foreign exchange forecasting. Condensing from the three outcomes to the two implies that a lot of market noise has to be allocated in one way or the other to the influence of the up or down economic indicators, diluting the credibility of the latter so far as major up or down movements are concerned. More or less stable and minor fluctuations are in effect put on the same footing as significant movement. However, there may be other contexts where the economics would always indicate either a clear up or a clear down outcome, and in this case the two outcome model may be useful. The empirics reported below cover only the three outcome model.

2.3 Dynamic Persistence Models

Dynamic elements can appear in the above models via lagged values of dependent variables. Thus if the objective is a directional forecast of exchange rates over the coming period, last period's actual value could appear among the explanatory variables **Z**.

Alternatively the dynamics can be more intrinsic to the model, and govern the way that the regimes themselves evolve. Recalling that these regimes describe directional changes, it could be that changes are persistent from one period to the next. Thus if the current direction is up, it is more likely that next period it will also be up. In such a persistence model, signal-output conditionals such as $P\ u|R_i$ and $P\ d|R_i$ remain constant, but the regimes themselves obey transition probabilities, of the form

$$P\ R_{i,t}\,|\,R_{j,t-1},\mathbf{Z}_t \ :\ P\ R_t \quad R_i\,|\,R_{t-1} \quad R_j,\mathbf{Z}_t\ , \tag{7}$$

where the symbol R_t is used to describe a regime-valued random process. The state probabilities evolve according to

$$P\ R_{i,t}\,|\,\mathbf{Z}^{\,t} \quad \sum_{j\ 1}^{3} P\ R_{i,t}\,|\,R_{j,t-1},\mathbf{Z}_t\ P\ R_{j,t-1}|\mathbf{Z}^{\,t-1} \tag{8}$$

starting from $P\ R_{i,0}\,|\,\mathbf{Z}_0$ as initial marginal distributions, where $i \quad 1,2,3$; $t \quad 1,2,$; and $\mathbf{Z}^{\,t}$ represent the history of the **Z** process up to time t. In some applications the transition probabilities might depend upon 'surprises' in the $\mathbf{Z}_t$, i.e. the part that could not have been predicted from $\mathbf{Z}^{\,t-1}$.

Expression (7) defines a Markov process, or a hidden Markov process (Baum and Petrie (1966)) where the regimes are obscured as in section 3. If the economy is now in a given state *R*, then it should most likely just stay there unless an economic event occurs. But if the exchange rate (for example) is currently in a stable 'no change' state and a very bad current account figure is announced, it should change to a down state with a greater probability. Thus the transition probabilities have to be made functions of the economic indicators **Z**, the latter now being reinterpreted as influences that will produce changes in state. The resulting Markov processes become nonhomogeneous.

At first sight, estimation of system (7) looks a formidable task, as there are now 9 transitional probability densities to specify and estimate. Note, however, that there are only 6 independent probabilities, as the columns of the transition matrix must sum to one. Further, the key entries are 'turning points' ($P\ R_{1,t}|R_{2,t-1},\mathbf{Z}_t$, $P\ R_{2,t}|R_{1,t-1},\mathbf{Z}_t$) and perhaps also 'breakout points' ($P\ R_{1,t}|R_{3,t-1},\mathbf{Z}_t$, $P\ R_{2,t}|R_{3,t-1},\mathbf{Z}_t$). This suggests that we might be able to get away with just 2-4 non homogeneous probability densities leaving the others constant, just as they are for standard homogeneous Markov models.

Alternatively, the regimes in the above could be reinterpreted as referring to levels rather than rates of change. Thus suppose the object to be forecasted was known to stay within a stable band over time. One could distinguish three levels: high, middle, low. A transition probability such as $P\ R_{i,t}|R_{j,t-1},\mathbf{Z}_t$ $(i \neq j)$, would now be interpreted as the probability of a transition from level *j* to level *i*.

3. OPERATIONAL MATTERS

In this section we look at some issues of regime delineation, model identification, and estimation procedures. In what follows, the forecasting objective concerns the change *Y* in an economic state variable of interest. Past values of *Y* can be observed but not precisely modelled. Instead the task is to estimate a categorical model governing the categorical regimes of change R_i into which the values of *Y* fall.

3.1 Regime Delineation

As earlier indicated, there is often value to adopting three outcomes, not just two, with a middle zone reserved for the absence of any significant up or down tendencies. However, for continuous variables this creates an operational problem of deciding boundaries between the up, stable or down outcomes. The boundaries may be treated as either known or unknown, and we shall consider the data densities that arise in each case.

(a) Known boundaries

This approach assumes that the investigator has established preassigned boundaries based on such considerations as a normal level of volatility, or what would consitute a break-out zone. With no real loss of generality, we could imagine that for a known number , the regimes are manifested by

$$\begin{aligned} &\text{Regime 1}: \quad Y \quad ; \\ &\text{Regime 2}: \quad Y \quad - \; ; \\ &\text{Regime 3}: \; - \quad Y \quad . \end{aligned}$$

The index functions associated with these sets will be denoted as $r_i\ Y,$. For example, $r_1\ Y,$ 1 if and only if *Y* ; 0, otherwise. In the above, the observable *Y* plays the role of an observable signalling variable, just as it does in probit analysis.

The likelihood element associated with an observation *Y* is a straightforward generalisation to three regimes of the standard probit model. It can be expressed in compact form as:

$$l_R \;\; Y, \mathbf{Z}; \;\; , \qquad \sum_{i\;\;1}^{3} r_i \;\; Y, \quad P \;\; R_i|\mathbf{Z}, \quad . \tag{9}$$

Although it contains Y as an observable argument, the likelihood element (9) refers to regime-valued events, emphasised by means of the subscript R. The effect is to single out the regime R_i that is actually observed. Thus if regime $i \quad 1$ is observed then $r_1 \;\; Y, \qquad 1$, while $r_2 \;\; Y, \qquad r_3 \;\; Y, \qquad 0$. We end up with $p \;\; R_1|\mathbf{Z}, \quad$ as the data density element that applies for this particular period. The classic probit likelihood function is precisely of this form. It has just two regimes with probabilities $p \;\; R_1|\mathbf{Z}, \quad$ and $p \;\; R_2|\mathbf{Z}, \qquad 1 - p \;\; R_1|\mathbf{Z},$ and the likelihood function over $t \quad 1, 2, \ldots T$ is divided into those periods where regime 1 and regime 2 apply. In both models, expression (9) could be described as a likelihood element generator function.

(b) Unknown boundaries

A more interesting approach is to model the regime boundaries and membership functions as fuzzy in nature. The original idea (Zadeh (1965); Zadeh and Bellman (1970)) was that the investigator might have some idea of where the boundaries might be, but less so as any more precise placement. One can assign values between zero and unity to the *degree* of membership of each regime, so that a given observation Y is not allocated absolutely to just one or the other.

An alternative interpretation in the present context can be constructed from more classical probability ideas. Here the boundary markers in any given period are sharp, but they can vary across different time periods, simply because some periods are naturally more volatile than others. We cannot observe the boundaries for any single period, but we can hope to model and estimate their probability distribution. The latter then effectively define the fuzzy membership functions. Put this way, it becomes clear that the regime boundaries are in the nature of hidden variables, or incomplete data, and the EM method potentially applies to the subsequent maximum likelihood solution for the model as a whole. What follows describes this approach.

To link with the known boundary case, suppose that we replace the firm boundary marker parameter with a random variable , and let us assume that $\sim N \;\; , \;\; ^2$. It is also assumed that is statistically independent of Y and $\mathbf{Z}$.; the boundary markers are simply noise parameters. The regime index functions now become of the form $r_i \;\; Y, \quad$, where the parameters , are suppressed for brevity. For instance, $r_1 \;\; Y, \qquad 1$ if and only if $Y \qquad$, so $R \quad R_1$ in this case and regime 1 applies. Also, for any given value of Y,

$$E \quad r_1 \;\; Y, \qquad \int_{-}^{Y} n \quad ; \;\; , \;\; ^2 \, d \qquad Y; \;\; , \;\; ^2 \, , \tag{10}$$

where $n \quad ; \;\; , \;\; ^2$ denotes the normal density function and $\quad Y; \;\; , \;\; ^2$ is the corresponding distribution function. Note that $\quad Y; \;\; , \;\; ^2$ has no reference to any statement about the data generating process of Y itself; the underlying probability distribution is that of , which is independent of Y.

Similarly, take expectations of the other two index functions $E \quad r_2 \;\; Y, \quad$ and $E \quad r_3 \;\; Y, \quad$, and let $_i \;\; Y; \;\; , \quad$ denote the respective results. We end up with

$$\begin{aligned} {}_1 \;\; Y; \;\; , \qquad & \quad Y; \;\; , \;\; ^2 \, ; \\ {}_2 \;\; Y; \;\; , \qquad & 1 - \quad Y; - \;\; , \;\; ^2 \, ; \\ {}_3 \;\; Y; \;\; , \qquad & 1 - {}_1 \;\; Y; \;\; , \quad - {}_2 \;\; Y; \;\; , \quad . \end{aligned} \tag{11}$$

Figure 3 plots these functions, which we can call the *fuzzy membership functions*. They are mutually symmetric, centered at zero, though other forms could be devised that are not necessarily symmetric, e.g. because the upper and lower boundary markers are not symmetric about zero. Note that for all Y, $\ _i\ Y\ \geq 0$, and $\sum_{i\ 1}^{3}\ _i\quad 1$.There is a width or location parameter ( ) and a sharpness parameter ( ), which could potentially be estimated along with the substantive model parameters .

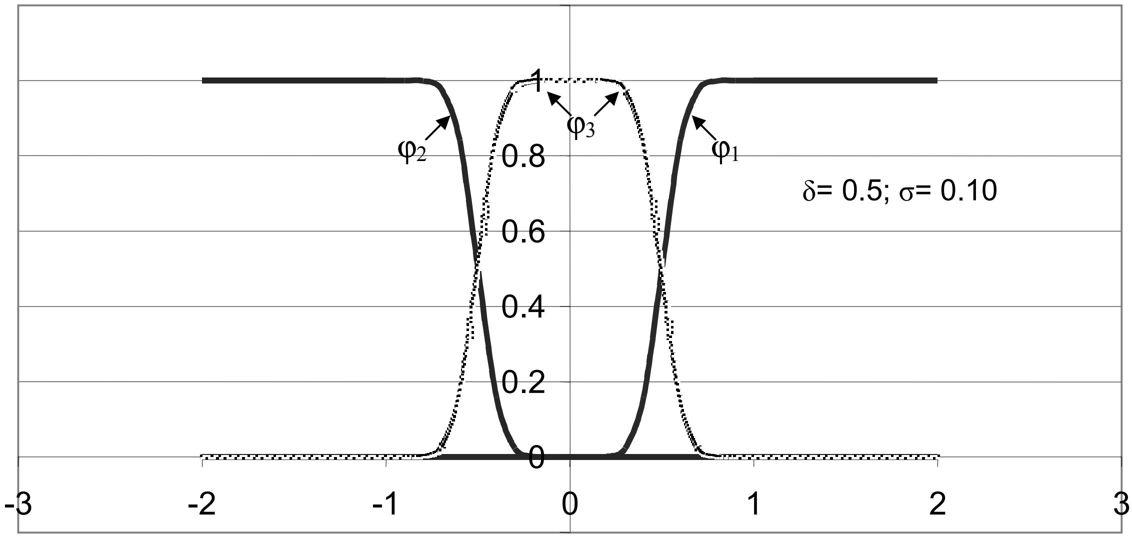

Figure 3: **Fuzzy Membership Functions**, **Normal Analogues**

The boundary markers (- , cannot be observed, so that observations Y cannot be allocated with certainty into one or other of the regimes R. However, we can obtain a *fuzzy likelihood element* by replacing the exact membership function $r_i\ Y,$ in expression (9) by the membership functions $\ _i\ Y;\ ,$ to obtain:

$$\bar{l}_R\ Y,\mathbf{Z};\ ,\ ,\qquad \sum_{i\ 1}^{3}\ {}_i\ Y;\ ,\quad P\ R_i|\mathbf{Z}, \tag{12}$$

Comparing with expression (9) for the known boundary case, it is easy to demonstrate that $\bar{l}_R\ Y,\mathbf{Z};\ ,\ ,\qquad E\ \ l_R\ Y,\mathbf{Z},\ ;\ $, motivating the notation. The effect of expression (12) is to give probability mass to all the regimes rather than just the one, to a degree depending on the strength of membership of observation Y in the respective regimes. It can be viewed as the marginal probability of regime-observation combinations, once the unobservable boundary uncertainty has been integrated out. Operationally, the device of replacing the unobservable index variables $r_i\ Y,$ by their expected values $\ _i\ Y;\ ,$ corresponds to the EM methodology, wherein hidden or incomplete data are replaced by their expectations, conditional upon what data is available.

The quasi likelihood element (12) will apply whether or not the boundary parameters , have been specified in advance. If either or both , are unknown, they become parameters to be estimated along with the substantive model parameters .

3.2 Regime Differentiation

It may be of interest to test whether the regimes are differentiated in terms of their connections, causal or otherwise, with the different economic indicators. Suppose, for instance, that there was really no need to differentiate between the regimes, e.g. because a simple linear model was operative between the **Z** variables and the output variable to be forecasted. Within the framework of model (1) above, this would correspond to a nesting where $_d \; - \; _u$ and a correlation between $_u$ and $_d$ of -1. The model becomes automatically complete and there are only two regimes possible, up or down. Alternatively, setting $_d \; - \; _u$ but continuing to assume $_u$ and $_d$ uncorrelated would drive an uncertainty wedge between two basically identical regimes, allowing for a stable or no-change band in between, a 'threshold for change' type of effect. At a minimum, we suggest testing for parameter commonality by first consolidating the **Z**'s to a common set and then testing $_d \; - \; _u$, or equivalently setting $\mathbf{Z}_d \; -\mathbf{Z}_d$ and testing $_d \quad _u$. Likelihood ratio tests can be used on all or some of the parameter pairs.

3.3 Estimation Procedures

Estimation procedures described in this section are based on maximum likelihood (ML). As previously noted, if the boundary markers are not precisely known, it has more of the character of the EM variant of ML. In addition, it may be necessary to use overlapping data, especially for macroeconomic work. We encompass such extensions as quasi maximum likelihood (QML) and comment below on implications.

For the static model, the quasi likelihood function is simply the product of the one-period elements as given by expression (12):

$$L_R \quad Y,\mathbf{Z} \quad ^T; \; , \; , \qquad \prod_{t \; 1}^{T} l_R \; Y_t, \mathbf{Z}_t; \; , \; , \tag{13}$$

where $Y,\mathbf{Z} \quad ^T$ denotes the history of observations from $t \quad 1,2,\ldots T$. Expression (13) refers to the more general case where neither of the boundary parameters $\;,\;$ are known; otherwise these parameters can be suppressed.

The more general QML for the dynamic persistence model (8) can be established as follows. Assume that the boundary markers $_t$ are serially independent, so that all intertemporal information is passed via the regimes R_{it}. Define a diagonal matrix of the membership functions:

$$_t \; , \quad : \begin{bmatrix} _1 \; Y_t; \; , & 0 & 0 \\ 0 & _2 \; Y_t; \; , & 0 \\ 0 & 0 & _3 \; Y_t; \; , \end{bmatrix}.$$

This can be combined with a transition matrix defined as

$$_t \quad : \begin{bmatrix} P \; R_{1t}|R_{1t-1},\mathbf{Z}_t; & P \; R_{2t}|R_{1t-1},\mathbf{Z}_t; & P \; R_{3t}|R_{1t-1},\mathbf{Z}_t; \\ P \; R_{1t}|R_{2t-1},\mathbf{Z}_t; & P \; R_{2t}|R_{2t-1},\mathbf{Z}_t; & P \; R_{3t}|R_{2t-1},\mathbf{Z}_t; \\ P \; R_{1t}|R_{3t-1},\mathbf{Z}_t; & P \; R_{2t}|R_{3t-1},\mathbf{Z}_t; & P \; R_{3t}|R_{3t-1},\mathbf{Z}_t; \end{bmatrix}$$

The intended QML can then be given as

$$L_R \quad Y,\mathbf{Z} \quad ^T; \; , \; , \qquad _1 \; ' \left(\prod_{t \; 1}^{T} \quad _t \quad _t \; , \right) _3, \tag{14}$$

where $_1$ is a stationary distribution derived from $_1$, and $_3$ is a $3 \quad 1$ vector of ones. The quasi-likelihood function (14) is similar to the likelihood function of a standard hidden Markov model. The only difference is that the transition matrix is a function of other explanatory variables, **Z**. If each row is the same, then the QML function for each period is identical to (13), so that the quasi-likelihood function of all observations can be written as a product of one-period quasi-likelihood functions. We adopt the static form (13) for the empirical data analysis reported below. In this work, the regime 'smearing' parameter will be preset as a known number, say $\bar{}$ while the position parameter will be estimated along with the parameters.The probability elements $P\ R_{it}|\mathbf{Z}_t;$ will be specified by the normal distribution function as in (1) and (5). The parameter estimates are collectively given by:

$$\hat{}_T : \quad \arg\max_{,} \quad T^{-1} \sum_{t\ 1}^{T} \ln \bar{l}_R\ Y_t, \mathbf{Z}_t;\ ,\ , \bar{} \quad , \tag{15}$$

where $\hat{}_T$ stands for $\hat{}_T', \hat{}_T\ '$. For brevity in what follows, we write $T^{-1} \sum_{t\ 1}^{T} \bar{l}_R\ Y_t, \mathbf{Z}_t;\ ,\ , \bar{} \quad \bar{L}_T \quad ; \bar{}$.

As earlier mentioned, the quasi-likelihood function is not necessarily the likelihood function corresponding to the true underlying data generation process. First, the conditional regime likelihood function, given $\mathbf{Z}_t$, can be different from $\bar{l}_R\ Y_t, \mathbf{Z}_t;\ ,\ , \bar{}$. If so, the information matrix inequality does not hold. Secondly, the given QML function ignores any dynamic persistence, and possibly leads to inefficient QMLE and an incorrect information matrix or efficiency bounds. On the other hand, it can be valid in this respect if all serial correlation can be captured among the predetermined explanatory variables, $\mathbf{Z}_t$. Thirdly, the fuzziness in regime boundaries implies an EM type formulation.

Nevertheless, the given likelihood function is flexible in the sense that it is specified by focusing on available economic information, and some of the technical difficulties arsing from model incompleteness can be resolved. Under suitable conditions, we can claim that $\hat{}_T$ as given by (15) has the almost sure limit $\quad \arg\max\ E\ \bar{L}_T \quad ; \bar{}$ (see Assumption I, Appendix A). In such a formulation, the quasi-likelihood function and convergence limit do not necessarily conform with exactly how the data are generated. The limit can be defined as the parameter explaining how closely the data would be generated according to the hypothesized model in terms of likelihood.

Specifying a quasi-likelihood function gives rise to a different asymptotic distribution from standard maximum likelihood estimators. We can still claim that the QMLE obeys asymptotic normality. Under Assumptions I and II in the Appendix A, it follows that

$$\sqrt{T}\ \hat{}_T - \quad \overset{A}{} N\ 0,\ -A \quad ^{-1} B \quad -A \quad ^{-1} , \tag{16}$$

where $A \quad : \quad E \quad ^2 \bar{L}_T \quad ; \bar{}$, $B \quad : \text{ avar } \sqrt{T} \bar{L}_T \quad ; \bar{}$, and 'avar' stands for the asymptotic variance of the given argument. Note that $-A$ is not necessarily the same as B , which should hold if the likelihood function is correctly specified. We estimate A by applying the strong uniform law of large numbers to $T^{-1} \bar{L}_T \ \hat{}_T; \bar{}$ and the convergence in probability of $\hat{}_T$ to . To estimate B we rely on Newey and West's (1987) method.

Experience thus far is that identification or maximisation difficulties can arise in the case where regime boundaries have to be estimated. For instance if the smearing parameter becomes large, the regimes become indistinguishable, the more so if the parameters are poorly identified. This is why a pre-set $\bar{}$ has been used in the empirical work reported below. Likewise, the parameter must be restricted to be positive and if allowed to become too large, all the data will be attributed to just the one regime, namely the stable zone. Operationally, one looks to the existence of an interior maximum within an interval that will attribute mass from all three regimes to the sample observations.

4. EMPIRICAL APPLICATION

The selected application is to hedging the corporate terms of trade of NZ dairy exporters, a profitability index taken as the ratio of product prices to input prices, both top and bottom denominated in terms of the NZ dollar (NZD). Although domestic inflation and commodity prices play a role, the most important single determinant is the USD versus NZD exchange rate, as most dairy exports contracts are denominated in the USD. The exchange rate fluctuates in a wide band over time, even if the band itself has not shifted very much up or down. A decision not to hedge the currency exposure leads to episodes where the industry is squeezed between a high NZD and low commodity prices. Welfare can usually, but not always, be improved by selling the USD receipts forward, for there is a chronic forward discount of the NZD associated with higher NZ interest rates. A fuller account of the context may be found in Bowden and Zhu (2006). The latter paper establishes forwards-based hedge rules of the passive or smoothing type, where the proportion of forwards used is invariant over time. In the present application we explore the effectiveness of a more active hedging policy, based on using current economic data to forecast the exchange rate for the coming period. The hedge ratio will therefore continually change over time, a conditional rather then unconditional hedge.

4.1 Forecasting

The efficient capital markets hypothesis would say that the only informative forecast, based on publicly available data, is the forward rate. However there is a large and variable forward rate discount on the NZD, and the forward rate can over some intervals be negatively correlated with the end of period spot rate. Indeed, the poor forecasting performance of the forward rate is well known in the foreign exchange literature (e.g. Hodrick 1987 for a review). On the other hand, Bowden (2004) has noted a close connection between the NZD exchange rate and the housing market, in the first instance created by the inflow of offshore funding for mortgages (the ‘hoovering effect’), with a further influence on the balance of payments capital account being provided by the relative fortunes of the NZ and US stock markets and business cycle. Adjustments in the exchange rate tend to follow major movements in these variables. They also often occur at times when the exchange rate is already at historic highs or lows, suggesting error correction elements in the level of the exchange rate relative to a historical equlibrium level, taken as constant. Sharp corrections are also seen as more likely when the balance of payments is seen as weak, e.g. as the build up of large current account imbalances. Unlike the US dollar, the NZ dollar is not seen as a major reserve currency! Taken together, the above suggests that exchange rate forecasting can be episodically successful, based on unusual movements or exposures, and provided one limits ambitions in the first instance to picking the direction of the movement, rather than necessarily the new equilibrium value. Table 1 summarizes the variables used in what follows and their timing conventions.

Table 1a: **Economic Variables Used for Yearly Forecasting**

Variables	Definition and Timeing Conventions
Exchange rate	$\ln NZD/USD_{t\ 4} - \ln NZD/USD_{t}$
Current account balance as % GDP	CAB/GDP_{t-8}
House price annual change index	$HPI_t - HPI_{t-4}\ /HPI_{t-4}$
Exchange rate as smoothed level	$\frac{1}{8}\sum_{k\ 0}^{7} \ln NZD/USD_{t-k}$
Relative share price index NZ/US	$\ln SPI_{NZ}/SPI_{US\ t} - \ln SPI_{NZ}/SPI_{US\ t-4}$
Relative GDP ratio index NZ/US	$\ln GDP_{NZ}/GDP_{US\ t} - \ln GDP_{NZ}/GDP_{US\ t-4}$

Table 1b: **Economic Variables Used for Quarterly Forecasting**

Variables	Definition and Timeing Conventions
Exchange rate	$\ln NZD/USD_{t\ 1} - \ln NZD/USD_{t}$
Commodity price index	$\ln CMP_{t} - \ln CMP_{t-1}$
House price quarterly change index	$HPI_{t-3} - HPI_{t-4}\ /HPI_{t-4}$
Relative share price index NZ/US	$\ln SPI_{NZ}/SPI_{US\ t} - \ln SPI_{NZ}/SPI_{US\ t-1}$
Relative GDP ratio index NZ/US	$\ln GDP_{NZ}/GDP_{US\ t} - \ln GDP_{NZ}/GDP_{US\ t-1}$

The endogenous variable is the NZD/USD exchange rate, measured with the NZD as the commodity currency (e.g. $1NZD \quad 0.7123USD$). The regime model is taken as static, meaning that changes in the exchange rate are assumed not to have persistence, apart from that derived from the **Z** variables. To obtain regime probabilities, the normal version of the model was employed. Three fuzzy outcomes were employed, corresponding to up, down and stable. Allowing both the position and spread as free fuzzy parameters led to indications that the spread parameter was poorly identified relative to the beta parameters. Based on a priori assessments as to regions of doubt for the boundaries, the fuzzy spread parameter was pre-set as 0.001, leaving the mean boundary delimiter to be estimated as the empirical regime marker of primary interest. The estimation method was quasi-maximum likelihood based on expression (15) using the computational routine 'SQPSOLVE' from Gauss version 5.02. The data is quarterly, with 62 observations from Q3 1989 to Q4, 2004. Sources are listed in the Appendix B. Official GDP and house price data is available only quarterly. This meant that in order to focus on annual changes, use of overlapping data became necessary, though not in the one-quarter based forecasting. Potential inefficiencies or biases due to the use of overlapping data have not been explored. Year on year changes are often used in the exogenous variables to improve the signal to noise ratio.

Criteria for inclusion or exclusion of indicator variables were based primarily on asymptotic *t* ratios and likelihood ratio tests from the QML estimation phase, and secondarily on contributions to the consequential hedging performance. Table 2 gives the estimated beta values and their asymptotic *t* ratios for the up and down indicators as listed and defined in table 1. We retained three significant up indicators, namely house prices, relative share prices and GDP ratios. Two down indicators were retained. The smoothed level has a primary impact when high, indicating the riskiness of exposure to the NZD at higher levels. The current account balance is important for similar reasons.

Table 2: **Directional Indicators and Parameter Estimates - Annual Forecasting Horizon**

Variables	Role	Estimates	t ratios
House price annual change index	up	11.522	2.940
Relative share price index NZ/US	up	9.049	2.815
Relative GDP ratio index NZ/US	up	45.278	3.186
Exchange rate as smoothed level	down	1.873	2.233
Current account balance as % GDP	down	-28.749	-2.609
Fuzzy boundary (		0.043	21.364

NOTE: BIC 90.282, and LR statistic testing all zero coefficients 59.473.

In addition we applied a likelihood ratio (LR) test as to whether all the indicator coefficients ( ) are zero, so that they provide no predictive information. Under the null, it follows from expressions (1) and (5) that $P\ R_{1t}|\mathbf{Z}_t;$ $P\ R_{2t}|\mathbf{Z}_t;$ $P\ R_{3t}|\mathbf{Z}_t;$ 1/3, so that estimating the null model is trivial. The resulting LR statistic decisively rejects the null hypothesis that they are zero.

Finally, the regimes are indeed quite different. Consolidating the **Z**'s as suggested in Section 3.2, followed by $_d$ $-$ $_u$, was rejected at 5% significance level, with all beta coefficients significantly different. This excludes alternative models such as simple linearity.

Choice of the indicator variables can depend on the length of the forecasting horizon. Table 3 is based on shorter term forecasting just one quarter ahead. Evidently, short run noise is obscuring the effect of the economic variables to a greater extent than with the longer forecasting horizon. Computed t ratios are not so significant as in the annual data case.

Table 3: **Directional Indicators and Parameter Estimates - Quarterly Forecasting Horizon**

Variables	Role	Estimates	t ratios
House price quarterly change index	up	25.268	2.087
Commodity price index	up	4.677	0.770
Relative share price index NZ/US	down	-4.416	-1.769
Relative GDP ratio index NZ/US	down	-35.134	-2.143
Fuzzy boundary (		0.005	34.972

NOTE: BIC 134, and LR statistic testing all zero coefficients 19.17.

4.2 Use in Hedging

At any time t, we can imagine a corporate treasury manager estimating the regime probabilities over the forecast horizon, as outlined above. These are to be used as an input into the hedge ratios that will protect the firm against an adverse outcome, in this case of the exchange rate, over the coming horizon. Hedging is to be done (we shall imagine) by buying or selling the foreign currency forward. Let h_t be the hedge ratio as a decision variable, meaning the proportion of the spot exposure that is to be protected. It is assumed that treasury policy requires $0 \leq h_t \leq 1$. The case h_t 1 would correspond to complete forward cover. As earlier mentioned, this is good steady state policy in the chosen context, because it takes advantage of the usual forward discount on the Kiwi or premium on the USD. However, on occasion it has been quite the wrong thing to do, with unprotected spot as much the better choice (see below). The issue is whether we can design a better hedge policy by allowing h_t to be variable, constituting an active rather than passive hedge rule.

At current time t suppose that the estimated regime probabilities for the coming forecast period (e.g. $t+1$ or $t+4$) are

$$\hat{p}_{jt+1} = P(R_{jt+1}|\mathbf{Z}_t, \hat{\ })\tag{17}$$

$j = 1,2,3$. Similarly with $t+4$ for one year ahead forecasting. The basic problem is how to map these forecast regime probabilities into the desired hedge ratio h_t. There are a number of ways of doing this.

One method is to fix as parameters q_u and q_d say, the size of the up and down jumps or ticks, and use these together with the regime probabilities to mimic a trinomial process, setting the R_3 tick as zero. The tick values q_u, q_d can be estimated by a historical least squares fit of the generated series with the actual. Using the tick values and the estimated up and down probabilities, one can estimate the expected value of the coming spot rate and adjust the hedge ratio according to the difference between the expected future spot rate and the currently quoted forward rate. One would expect this sort of technique to work better for very short horizon hedging, where the trinomial tick process provides a better approximation.

Better results, however, can be obtained by making fuller optimizing use of the three possible outcome regimes, in conjunction with other information such as whether the exchange rate is currently high or low relative to history. Thus a more effective rule - for this particular context - is based on the observation that historical values of the NZD/USD exchange rate have fluctuated within a broad band, but one without any discernible trend. Suppose we divide the historical series into three zones: high, middle and low. The consequences for NZ exporters of the NZD/USD exchange rate moving up are worse if the exchange rate itself is already high. On the other hand, if the NZD/USD is at a historical low, exporters would be less troubled by the prospect of a further up movement, or even if the rate stayed the same. Modifying this is the chronic forward rate discount of the NZD which would lead to a bias in favor of using the forward rate. One could imagine a 3×3 matrix of loading weights $_{ij}$ proportional to the marginal utilities of movement up, down or stable (j), given that the current NZD/USD rate is at a historical high, medium or low level (i). For instance, $_{13}$ would be the loading if the current state was in the high historical zone and the movement was stable.

The hedge rule would then be of the form

$$h_t = h_{i\,t} = \sum_{j=1}^{3} {}_{ij}\hat{p}_{jt+1}.\tag{18}$$

If at time t one observes that the current exchange rate is in historical zone i, then one weights the estimated direction probabilities with the loadings appropriate for zone i. The hedge ratio is the expected value of the loadings. The weights, collectively , can be determined by maximizing a chosen welfare function, assuming that rule (18) had been applied historically.

In the present application the hedge weights are chosen to maximize a conditional value at risk type utility function. The conditional value at risk (CVaR) utility function (Bowden and Zhu 2006) is equivalent to assuming that an otherwise risk neutral manager has been compelled to write put options in the sensitive zone (see Appendix C) . The dependent variable is the farmer terms of trade, constructed as the NZD earnings to farmers from export sales in USD, divided by an input price index. Taking the log of this is a proxy for the net farm profit margin, a standard accounting metric for profitability. Thus if the farmer terms of trade weakens beyond a certain point, the losses become especially severe. In order to economize slightly on the number of loading parameters to be estimated, it is assumed that the elements of the last row of are all equal. The last row refers to the good state for NZ exporters (low NZD), so one can assume a natural tendency to simply use the forward rate no matter what the outcome is. The loadings to be estimated are constrained to the range $0 \leq {}_{ij} \leq 1$, which will similarly constrain the hedge ratio.

Table 4 shows that the optimized loadings are mostly either unity or zero. The differences are most marked in the high zone, where the unhedged exporter is suffering from the high NZ dollar. If the direction call is up, then the exporter would definitely want complete protection; but if down, then the exporter would want to be fully exposed to the spot rate and forego the forward altogether.

Table 4: **Optimised Hedge Loadings - Elements of**

Current state *Movement*	up	down	stable
High band spot rate	1	0	99.96%
Middle band spot rate	1	44.17%	1
Lower band spot rate	1	1	1

Once the hedge ratios are ascertained, we can compute the effective conversion exchange rate for the exports, made up of the spot rate and the forward rate in proportions indicated by the hedge ratio for each time period. In turn, this can be used to calculate the farmer's hedged terms of trade, on a historical basis. This can be compared with the historical terms of trade without hedging, and also with that derived from using just the forward rate to make the conversion. Table 5 compares the outcomes using a number of common metrics. The value at risk (VaR) is the lower 10% critical point for the marginal distribution of the terms of trade, as though they all came from a common underlying distribution, while the CVaR is the conditional expected value given that the terms of trade is censored to be less than the VaR critical point - it is a measure of the mass in the left hand tail. The optimized hedge is superior on all measures to either remaining unhedged or fully hedging with the forward.

Table 5: **Effective Conversion Exchange Rate - Some Statistics**

	Unhedged case	One year forward hedge	Unconditional optimum	Conditional optimum
Mean	-1.685	-1.637	-1.660	-1.648
Variance	0.023	0.015	0.015	0.013
10% VaR	-1.858	-1.803	-1.805	-1.797
10% CVaR	-1.884	-1.828	-1.827	-1.817
E U	0.091	0.190	0.197	0.198

Figure 4 compares the history of the farmer terms of trade under the three alternatives. The optimized outcome generally tracks closely the simple forward as the 100% hedge rule. However there is useful divergence over the interval 1997-99, which saw the NZD crumble in the aftermath of the Asian crisis. It was a bad idea to follow a simple forward strategy at this time, which locked in at the high NZD exchange rate prior to the crash. The optimized strategy uses the economic information current at the time and elects to remain with the unhedged spot rate. Note that both the pure forward and optimized strategies managed to avoid the adverse effects of the high NZD from 2002 onwards.

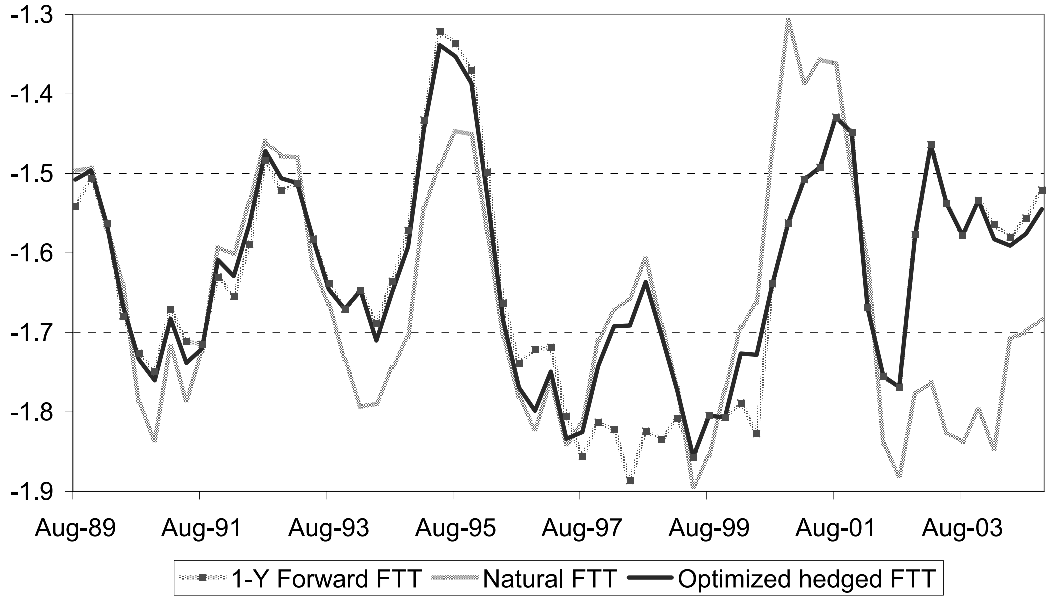

Figure 4: **Historical Comparison**, **One-Year Hedging Horizon**

5. CONCLUDING REMARKS

Information management and use in real world forecasting and risk management can be active or passive, or somewhere in between. At one extreme, the manager knows enough about the economy and the markets to be able to formulate a complete and informative economic model, one capable of accurate value forecasts. In econometric terms, one would think of this as a full information approach, and the estimation would be based on a likelihood function derived from a well specified model of the underlying data generation process. At the other extreme, a manager would use only implicit market forecasts or certainty equivalents, arguing that given market efficiency, he or she should not even try to do better. This is a passive view of informational gathering and use. But it is worth remembering that it depends itself on the kind of models, mental or otherwise, used by market practitioners. Unfortunately, there is empirical evidence that market outcomes do not always square up to the precepts of economic theory and informational efficiency. In the chosen context of the present paper, namely exchange rate forecasting, the theoretical result from martingale pricing in complete markets is that the forward rate is an optimal predictor, with or without risk aversion. But only the most passive and rule-driven manager would rely on the predictive content of the forward rate, especially for a small open economy. It is natural to try and do a little better by using what information and economic insight might be available, while perhaps recognizing the potentially transitory nature of the advantage to be gained until the market at large catches on. A framework of incomplete directional forecasting is adapted for this sort of exigency. It outperforms passive forecasting in the particular context of the paper, founded on historical observation and performance. This is not to say that the forecasting expressions that result will continue to do well into the future. But this does not limit the usefulness of the techniques themselves as a methodological window into the problem of forecasting with limited-information.

Limited information, especially as to the nature of the underlying data generation process, does place some special demands on the econometrics and on the use to be made of it. Neuronal net modelling has been designed just for such situations. However, fresh informational problems arise in circumstances where data limitations preclude using the full power of artificial learning algorithms or procedures, as is commonly the case where macroeconomic data is involved. It becomes necessary to introduce additional a priori econometric structure into the problem, supplemented by whatever economic theory seems realistic and feasible. A further problem is that the true underlying data generation process may not coincide precisely with the estimation model. In effect, one is replacing maximum likelihood by quasi-maximum likelihood, and convergence to an equivalent set of parameters under the specified model may not always be assured. Likewise an explicit methodology is necessary to relate hedge parameters to the resulting directional probabilities. But our finding here is that even very simple frameworks, involving only directional rather than complete value assessments, can produce a better result than informational passivity; or on the other hand investing large sums of time and money in large scale modelling efforts, with little prospect of any lasting success.

APPENDIX A: SUFFICIENT CONDITIONS FOR QUASI MAXIMUM LIKEHOOD ESTIMATION

Even if the true data generating porcess is unknown, limited information may nevertheless prove consistent with the data, in the sense that their parameter estimates (or quasi estimates) converge. The following are some general assumptions that will suffice, together with comments relating to the context of the present paper.

Assumptions I

(i) $Y_t, \mathbf{Z}_t'\,'$: t 1,2, is a set of strictly stationary and ergodic processes;

(ii) the space Θ for is compact in $\mathbb{R}^p$ (p $\mathbb{N}$);

(iii) for a stationary, ergodic, and integrable D_t, sup $_\Theta |\ln \bar{l}_{R,t}$;$^-$ $| \leq D_t$;

(iv) there is a unique maximizer of $E\,\bar{L}_T$;$^-$ in the interior part of Θ.

Comments:

(a) Assumption I(i) describes the conditions for the data generating process of the data. The stationarity and ergodicity condition is crucial in applying the asymptotic theory.

(b) Assumptions I(iii, iv) specify the conditions for the consistence of the QMLE, $\hat{}_n$. By Assumption I(iii), the strong uniform law of large numbers holds for $n^{-1} \sum \ln \bar{l}_{R,t}$;$^-$, so that $\hat{}_n$ converges a.s. to given that $E\,\bar{L}_T$;$^-$ is identified.

Assumptions II

(i) $\ln \bar{l}_{R,t}$;$^-$, $_t$ is an adapted mixingale of size -1 (McLeish, 1974), where $_t$ is a smallest -algebra generated by $y_1, \mathbf{z}_1, \ldots, y_t, \mathbf{z}_t$;

(ii) E $\ln \bar{l}_{R,t}$;$^-$ $'$ $\ln \bar{l}_{R,t}$;$^-$;

(iii) B is positive definite, where avar is the asymptotic variance of given argument;

(iv) sup $_\Theta \|$ $^2 \ln \bar{l}_{R,t}$;$^-$ $\|$ D_t, where $\|$ $\|$ is the uniform metric;

(v) A : E $^2\bar{L}_T$;$^-$ is negative definite.

Comments:

(a) The central limit theorem (CLT) can be applied to $n^{-1/2} \sum$ $\ln \bar{l}_{R,t}$;$^-$)by Assumptions II(i to iii). Scott (1973) provides sufficient conditions for the central limit theorem, and White (1999, p. 125) proves the CLT given Assumptions II(i to iii).

(b) Assumptions II(iv and v) are used to approximate $\bar{L}_T$;$^-$ by a quadratic function. The standard second-order Taylor expansion can be applied to $\bar{L}_T$;$^-$.

(c) By Assumption II(iv), we can apply the strong uniform law of large numbers to the Hessian matrix. The negative definite Hessian matrix is necessary for a non-degenerate asymptotic distribution of the QMLE.

APPENDIX B: DATA DEFINITIONS AND SOURCES

Table 6 gives sources for the economic variables used in the paper.

Table 6: Data Definitions and Sources

Variables	Meanings	Sources
NZD	New Zealand dollar	Global Trade Information Services
USD	US dollar	Global Trade Information Services
HPI	House price index	Reserve Bank of New Zealand
SPI_{NZ}	NZ share price index	Morgan Stanley Capital International Inc.
SPI_{US}	US share price index	Morgan Stanley Capital International Inc.
CMP	NZ commodity price index (dairy, in USD)	ANZ Bank
GDP_{NZ}	NZ gross domestic product (current dollars)	Statistics NZ
GDP_{US}	US gross domestic product (current dollars)	International Monetary Fund
CAB	NZ current account balance as % GDP	Statistics NZ

APPENDIX C: THE CONDITIONAL VALUE AT RISK UTILITY FUNCTION

The utility function used for hedging has the basic form

$$U\ R;P \qquad R-P \qquad \min\ R-P,0$$

where the random variable *R* represents the variable to be hedged (in the present study the farmer log terms of trade, taken as the net profit margin). Taking the expected value over the min $\cdot$ function introduces the conditional value at risk element. Figure 5 illustrates the effect, which is though an otherwise risk neutral manager with a linear utility function AA has written put options on the value *R* with strike price P, so that the overall effect is CPA. An analogy is with corporate finance, where the threat of bankruptcy costs creates a third party claimant to corporate value (liquidators, lawyers etc) to be exercised in the event of collapse. The point *P* is chosen as a point of special sensitivity, often a desired value at risk point.

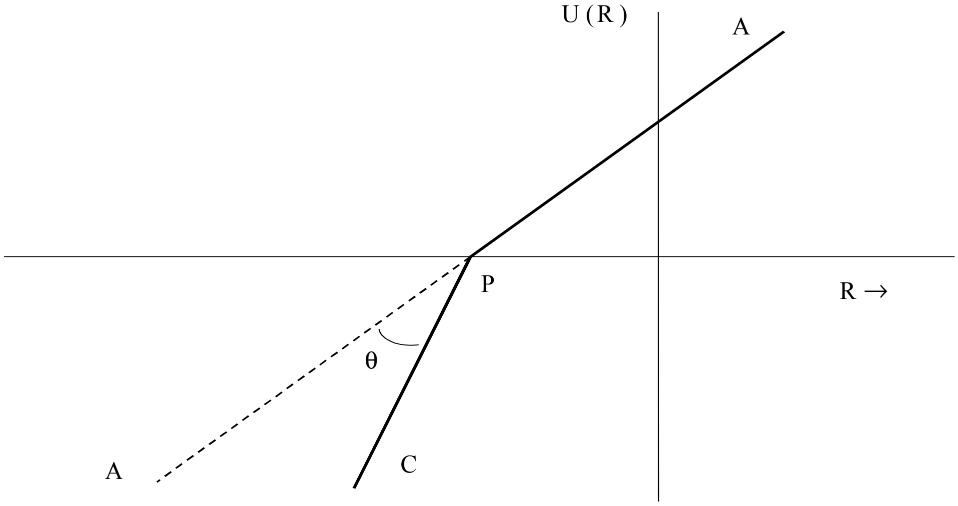

Figure 5: **Conditional Value at Risk Utility Function**

In the application of this paper, the sensitive point is set at P 1.8576, the idea being that if the terms of trade fall below figure, farmers will be in financial trouble with cash flows. If value at risk type critical points are used (say 10%), the parameter has to be set high enough to produce significant risk aversion, given the probability of R falling below P; we used 20. To improve numerical convergence properties with optimization, the strict stepwise function is replaced by a slightly smoothed version in the neighborhood of the point P. The smoothed version is itself derived from options theory as the Black Scholes price of the implied third party options mentioned above. For further details of the option-equivalent approach, see Bowden and Zhu (2006).

REFERENCES

Artzner, P., Delbaen, F., Eber, E.M., and Heath, D. (1997), "Thinking Coherently," *Risk*, 10, 68-71.

——————(1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 203-228.

Baum, L.E., and Petrie, T. (1966), "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," *Annals of Mathematical Satistics*, 37, 1554-1563.

Bickel, P.J., Ritov, Y., and T. Ryden (1998), "Asymptotic Normality of the Maximum-Likelihood Estimator for General Hidden Markov Models," *Annals of Statistics*, 26, 1614-37.

Bowden, R.J. (2004), "McHouse Prices, Capital Hoovering, and Real Exchange Rate Exposures," *New Zealand Economic Papers*, 38, 119-139.

————(2005), "Ordered Mean Difference Benchmarking, Utility Generators and Capital Market Equilibrium," *Journal of Business*, 78, 441-467.

Bowden, R.J., and Zhu, J. (2006), "Asymmetric Hedging of the Corporate Terms of Trade," *Journal of Futures Markets*, forthcoming.

Cappe, O., Moulines, E., and Ryden, T. (2005), *Inference in Hidden Markov Models*, New York: Springer Science Business Media.

Christoffersen, P. F., and Diebold, F. X. (2004), "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics," PIER Working Paper No. 04-009, University of Pennsylvania, The Penn Institute for Economic Research.

Dempster, A., Laird, N., and Rubin, D. (1977), "Maximum Likelihoood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, series B, 39, 1-38.

Diebold, F. X., and Mariano, R. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-265.

Ephraim, Y., and Merhav, N. (2002), "Hidden Markov Processes," *IEEE Transactions on Information Theory*, 48, 1518-1569.

Gradojevic, N., and Yang, W. (2000), "The Application of Artificial Neural Networks to Exchange Rate Forecasting: The Role of Market Microstructure Variables," Working Paper Series, 2000-23, Bank of Canada, Financial Markets Department.

Hartley, H. (1958), "Maximum Likelihood Estimation from Incomplete Data," *Biometrics*, 14, 174-94.

Hodrick, R. J. (1987), *The Empirical Evidence of the Efficiency of Forward and Futures Foreign Exchange Markets*, New York: Harwood Academic.

Kuan, C. M., and White, H. (1994), "Artificial Neural Networks: An Econometric Perspective," *Econometric Reviews*, 13, 1-91.

Leitch, G., and Tanner, J. E. (1991), "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures," *American Economic Review*, 81, 580-590.

Leung, M. T., Daouk, H., and Chen, A. S. (2000), "Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models," *International Journal of Forecasting*, 16, 173-190.

Levich, R. M. (2001), *International Financial Markets* (*2nd ed.*), NewYork: McGraw-Hill.

McCulloch, W.S., and Pitts, W. (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 5,

115-133.
McLeish, D.L. (1974), “Dependent Central Limit Theorems and Invariance Principles,” *Annals of Probability*, 2, 620-628.
Newey, W., and West, K. (1987), “A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703-708.
Pesaran, M. H., and Timmermann, A. (2004), “How Costly Is It to Ignore Breaks When Forecasting the Direction of a Time Series?,” *International Journal of Forecasting*, 20, 411-425.
Rosenblatt, R. (1957), “The Perceptron: A Perceiving and Recognizing Automaton,” Project PARA, Cornell Aeronautical Laboratory Report, 85-460-1.
————(1958), “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” *Psychological Reviews*, 62, 386-408.
Scott, D.C. (1973), “Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorohod Representation Approach,” *Advances in Applied Probability*, 6, 119-137.
Tkacz, G., and Hu, S. (1999), “Forecasting GDP Growth Using Artificial Neural Networks,” Working Papers, 99-3, Bank of Canada.
Turban, E. (1995), *Neural Networks Finance and Investment*: *Using Artificial Intelligence to Improve Real-World Performance* (2*nd ed*.), New York: McGraw-Hill.
White, H.L. (1999), *Asymptotic Theory for Econometricians* (2*nd ed*.), CA: Academic Press.
Zadeh, L.A. (1965), “Fuzzy Sets,” *Information and Control*, 8, 338-353.
Zadeh, L.A., and Bellman, R. (1970),“Decision-Making in a Fuzzy Environment,” *Management Science*, 17, 141-164.