How Costly is it to Ignore Breaks when Forecasting the Direction of a Time Series?*

M. Hashem Pesaran  Allan Timmermann
University of Cambridge  University of California, San Diego

November 9, 2001

Abstract

Much empirical evidence suggests that macroeconomic and financial time series are subject to occasional structural breaks. In this paper we present analytical results quantifying the effects of such breaks on the ability to forecast the sign or direction of a time-series that is subject to breaks. Our results suggest that it can be very costly to ignore breaks. Forecasting approaches that condition on the most recent break appear to produce significant improvements over unconditional approaches that use expanding or rolling estimation windows.

JEL Classifications: C22, G10.

Key Words: Sign prediction, estimation window, structural breaks.

*We thank James Chu, David Hendry, Adrian Pagan and seminar participants at Erasmus University, London School of Economics and University of Southern California for comments and discussion.
1. Introduction

In the most systematic study to date of a very large set of macroeconomic time series, Stock and Watson (1996) find evidence of structural breaks in the majority of the series they consider. Breaks or jumps in the parameters of economic forecasting models could arise due to a number of factors such as major changes in market sentiments, burst or creation of speculative bubes, regime switches in monetary and debt management (for example, from money supply targeting to inflation targeting or from short-term to long-term debt instruments). These possibilities are important because they fundamentally affect the extent to which economic variables are predictable. However, the implications of this evidence for directional forecasting have yet to be investigated.

In this short paper we characterize analytically the factors that determine the loss in directional forecasting accuracy from ignoring information about breaks. Unconditional methods for estimation of a forecasting model such as a rolling or an expanding window let the window size vary as a deterministic function of time. These methods will produce biased and inconsistent forecasts in the presence of breaks. A conditional approach that determines the window size on the basis of the estimated point of the most recent break can be expected to do better.

Instability in prediction models is particularly important for identifying turning points or the ‘direction’ of the market. Since the seminal work by Henriksson and Merton (1981) on market timing and predictability of the signs of security returns, there has been extensive interest in this problem in both economics and finance. Leitch and Tanner (1991) find that the correlation between a sign test and the profits made from following investment advice dominates the correlation between profits and standard statistical measures of prediction such as mean squared forecast error. Despite its importance, the problem of sign predictability when the underlying return generating process may have undergone a structural change has not yet been addressed in the forecasting literature. We consider this issue in the context of a simple linear regression model and compare both unconditional and conditional approaches to determination of the window size used in the estimation of a forecasting model.

The plan of the paper is as follows. Section 2 derives analytical results that allow us to quantify the market timing information in forecasting models that account for breaks relative to models that ignore these. Section 3 provides numerical examples
that illustrate the analytical results. Section 4 concludes and discusses the empirical relevance of our findings.

2. Sign Prediction under Breaks

In this section we explore analytically the effects of structural breaks on the directional forecasting accuracy of procedures that either ignore a break or account for it. We focus on directional forecasting accuracy since this is now an increasingly popular metric for forecasting performance. In contexts such as market timing this measure is closely related to the economic value of forecasts used in asset allocation decisions.

To keep the analysis tractable, we consider a linear regression model subject to a single structural break occurring at some date, \( T_1 \)

\[
\begin{align*}
y_t & = \beta'_1 x_t + u_t, \\
\beta'_2 x_t + u_t & \sim \text{IID}(0, \sigma^2_1), \quad t = 1, 2, \ldots, T_1 \\
\beta'_2 x_t + u_t & \sim \text{IID}(0, \sigma^2_2), \quad t = T_1 + 1, \ldots, T + 1. 
\end{align*}
\]

(1)

\( y_t \) is the variable that is being predicted, \( x_t \) is the \( p \times 1 \) vector of predictor variables, \( \beta'_i \) (\( i = 1, 2 \)) are \( p \times 1 \) vectors of regression coefficients, and \( u_t \) is a serially uncorrelated error term that is independently distributed of \( x_s \) for all \( t \) and \( s \), possibly with a shift in its variance from \( \sigma^2_1 \) to \( \sigma^2_2 \) at the time of the break point.

Let the forecast of \( y_{T+1} \) conditional on period \( T \) information be given by \( \hat{y}_{T+1} = \hat{\beta}' x_{T+1} \). The probability of correctly predicting the sign of \( y_{T+1} \) is given by \( \Pr(\hat{y}_{T+1} y_{T+1} > 0) \). To simplify the exposition, let \( p = 1 \), so that

\[
\begin{align*}
y_{T+1} & = \beta_2 x_{T+1} + u_{T+1}, \\
\hat{y}_{T+1} & = \beta_2 x_{T+1} + (\beta_1 - \beta_2) \theta_m x_{T+1} + x_{T+1} \left( \frac{\sum_{t=m}^{T} x_t u_t}{\sum_{t=m}^{T} x_t^2} \right) .
\end{align*}
\]

(2)

where \( \theta_m = \theta_m(T_1, T) = \frac{\sum_{t=m}^{T_1} x_t^2}{\sum_{t=m}^{T} x_t^2} \). We measure sign predictability by means of the nonparametric market timing test statistic of Pesaran and Timmermann (1992) which is asymptotically equivalent to the test developed originally by Henriksson and Merton (1981) but is more convenient to work with. Granger and Pesaran (2000) show that this market timing statistic can also be written as

\[
PT = \frac{\sqrt{n}(H - F)}{\left( \frac{\hat{\pi}(1 - \hat{\pi})}{\hat{\pi}(1 - \hat{\pi})} \right)^{1/2}},
\]

(3)
where $n$ is the number of observations in the forecast period, $H$ is the “hit rate” and $F$ is the “false alarm rate”, which are defined as

$$
H = \frac{\Pr(\hat{y}_{T+1} > 0, y_{T+1} > 0)}{\Pr(y_{T+1} > 0)},
$$

$$
F = \frac{\Pr(\hat{y}_{T+1} > 0, y_{T+1} < 0)}{\Pr(y_{T+1} < 0)}.
$$

Finally $\pi = \Pr(y_{T+1} > 0)$, and $\hat{\pi} = \Pr(\hat{y}_{T+1} > 0)$ are the probabilities that the realization and predicted values of returns are positive, respectively. The hit minus false alarm rate has a very intuitive interpretation as the probability of correctly predicting the sign of a positive return over the probability of wrongly predicting the sign of a negative return. It is equal to zero for all forecasts that do not have any information about the sign of returns so a necessary condition for market timing information is a strictly positive value of this statistic.

In general it is complicated to derive an analytical expression for the probability of correctly predicting the sign of excess returns. However, in the simple case where $u_t$ and $x_t$ are serially uncorrelated and normally distributed with $\sigma^2_1 = \sigma^2_2 = \sigma^2$, we can derive an expression that demonstrates how the probability of correctly predicting the sign of $y_{T+1}$ depends on whether or not a break is accounted for. Suppose the full data set consists of $v = v_1 + v_2$ observations where $v_1$ is the number of pre-break observations, while $v_2$ is the number of post-break observations. A conditional approach that uses only post-break information will estimate the forecasting model using $v_2$ data points, while the unconditional expanding window method uses all $v_1 + v_2$ observations.

To derive the market timing statistic we first state the joint distribution of the realized and predicted excess return. Suppose the state variable used in the prediction is normally distributed, $x_t \sim N(\mu_x, \omega^2)$. In this case the joint distribution is

$$
\begin{pmatrix}
y_{T+1} \\
\hat{y}_{T+1}
\end{pmatrix}
\sim IN\left\{\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, \Sigma\right\},
$$

where the means and covariance matrix are given by

$$
\begin{align*}
\mu_1 &= \beta_2 \mu_x, \\
\mu_2 &= \beta_2 \mu_x + \frac{(\beta_1 - \beta_2)\nu_1 \mu_x}{\nu}, \\
\Sigma &= \begin{pmatrix}
\beta_2^2 \omega^2 + \sigma^2 & g \\
g & h^2
\end{pmatrix}.
\end{align*}
$$
$g$ and $h$ are constants defined by

$$
h^2 \equiv V(\hat{y}_{T+1}) = \sigma^2 \kappa(\lambda, \delta, \nu) + \beta_2^2 \omega^2 +
\left(\beta_1 - \beta_2\right)^2 \omega^2 \left(\frac{\nu_1}{\nu}\right) \zeta + 2 \beta_2 (\beta_1 - \beta_2) \omega^2 \left(\frac{\nu_1}{\nu}\right),
$$

(7)

$$
g \equiv \text{Cov}(y_{T+1}, \hat{y}_{T+1}) = \beta_2^2 \omega^2 + \beta_2 (\beta_1 - \beta_2) \omega^2 (\nu_1/\nu),
$$

while $\kappa(\lambda, \delta, \nu)$ and $\zeta$ are constants that depend on the window size parameters and the means and variances of the underlying series as explained in the Appendix.

Some intuition can already be gathered from these expressions. Suppose that a window of $\nu_1 \geq 1$ pre-break observations has been used to estimate the regression coefficient of $x$. Then the estimated mean of excess returns, $\mu_2$, will be biased, although the forecast error variance will also be smaller than if only post-break observations were used ($\nu_1 = 0$). Likewise, the larger is the break size, $|\beta_1 - \beta_2|$, the weaker is the correlation between predicted and realized excess returns and thus the lower the sign test statistic.

The appendix proves the following result:

**Proposition 1** Suppose that the forecast error $u_t$ and the state variable $x_t$ are serially uncorrelated and normally distributed

$$
\begin{pmatrix}
u_t \\
x_t
\end{pmatrix} \sim \text{IN} \left\{ \begin{pmatrix} 0 \\ 0 \\ \sigma^2 \\ 0 \\ \omega^2 \end{pmatrix} \right\}.
$$

Then the hit minus false alarm rate (H-F) associated with the realizations and forecasts of returns $(y_{T+1}, \hat{y}_{T+1})$ is given by

$$
H - F = \frac{A_H}{\Phi(\tilde{\mu}_1)} - \frac{A_F}{1 - \Phi(\tilde{\mu}_1)},
$$

(8)

where $\tilde{\mu}_1 = \mu_1/\sqrt{\sigma^2 + \beta_2^2 \omega^2}$,

$$
\Phi(\tilde{\mu}_1) = \int_{-\infty}^{\tilde{\mu}_1} \left(2\pi\right)^{-1/2} \exp\left(-\frac{1}{2} a^2\right) da,
$$

$$
A_H = \text{Pr}(y_{T+1} > 0, \hat{y}_{T+1} > 0) = \int_{a_2 = -\mu_2}^{\infty} \int_{a_1 = -\mu_1}^{\infty} f(a_1, a_2) da_1 da_2,
$$

$^1$Notice that in the special case where $\mu_x = 0$, $\mu_1 = \mu_2 = 0$ and $\Phi(\mu_1/\sqrt{\beta_2^2 \omega^2 + \sigma^2}) = 1/2$, so $H - F = 0$. 

\[ A_F = \Pr(y_{T+1} < 0, \hat{y}_{T+1} > 0) = \int_{a_2 = -\mu_2}^{\infty} \int_{a_1 = -\infty}^{-\mu_1} f(a_1, a_2) da_1 da_2, \]

\[ f(a_1, a_2) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp(-\frac{1}{2} a' \Sigma^{-1} a), \]

\[ a = (a_1, a_2)'. \]

3. Numerical Results

The proposition in the previous section allows us to determine the exact value of the market timing statistic as a function of the window of data used to estimate the forecasting model and the size and timing of a break in the linear regression model.

Figure 1 uses the proposition to show how the market timing information in the predicted variable depends on whether an expanding or rolling window or the conditional break point method is used. The figure plots the hit minus false alarm rate as a function of the value of \( \beta_1 \) which tracks the break size \( |\beta_1 - \beta_2| \). The figure assumes that \( \beta_2 = 1, \mu_x = 0.5, \sigma = 6, \omega = 1.5 \) and \( v_1 = 100 \). In the context of a return forecasting model these parameters correspond to volatility of six percent and a population \( R^2 \)-value of 0.06. This matches empirical evidence on monthly US stock returns. The break in the regression coefficient ranges from 0 to 3 which is realistic in view of the substantial parameter variation found empirically for this type of data. The rolling window follows standard practice in economics and finance and uses 60 observations, while the expanding window uses the full set of \( v_1 = 100 \) pre-break observations in addition to the post-break data points \( v_2 \).

The upper window assumes that the break happened 10 observations ago \( (v_2 = 10) \). In the absence of a break, the expanding and rolling windows naturally produce better sign predictions than the post-break window which wrongly ignores pre-break data. However, as the break size increases to around one, the post-break window method begins to dominate. In the presence of a break, both the expanding and rolling windows produce biased forecasts since they include pre-break data to estimate the forecasting model. For break sizes above one, in fact these methods generate a negative sign statistic.

The middle window shows the case where the break happened 25 periods ago \( (v_2 = 25) \). If only post-break information is used, \( H - F \) is constant around 10 percent irrespective of break size, while the expanding window method gener-
ates negative values for a break size above 1.2 and the rolling window only does marginally better. As the break size $|\beta_1 - \beta_2|$ increases, the hit minus false alarm rate for the expanding and rolling windows continue to decline, suggesting that a forecasting model that fails to account for a break can lead to a severe deterioration in market timing performance.

The lower window in Figure 1 assumes that $v_1 = v_2 = 100$, implying that a break occurred 100 observations prior to the forecasting point. Since more post-break information is now available, the post-break hit minus false alarm rate rises to 11 percent. Although the rolling window does not use any pre-break data (the window size (60) is less than the time since the break (100)), it still underperforms slightly relative to the post-break model since it does not use all 100 post-break data points and hence is inefficient. However, for values of the break up to around 1, the three methods produce similar results. When the break size increases beyond this point, the performance of the expanding window that uses all pre-break data points quickly deteriorates.

4. Conclusion

Many economic and financial time series undergo sudden, large breaks reflecting institutional changes, regime switches or breakdowns in market mechanisms. Forecasting such series poses a difficult problem, particularly if one is interested in the sign of the variables as is frequently the case. In this paper we characterized analytically the factors that determine the forecasting performance of standard approaches to window selection when the true data generating process undergoes breaks.

Our results demonstrate the importance to directional forecasting of correctly selecting the window used to estimate the forecasting model. On the downside, a forecasting approach that conditions on a false breakpoint risks being inefficient as it does not use the full set of available data. On the upside, an approach that succeeds in correctly identifying a break can reduce the bias inherent in the rolling window and expanding window forecasts. When breaks are reasonably large and the distance to the most recent break is not too great, our results suggest that such gains can be very significant.

These findings are highly relevant to empirical forecasting. In Pesaran and Timmermann (2001) we find significant market timing gains from conditioning on
the estimated time of the most recent break point in an out-of-sample forecasting model of US stock returns. The results presented in this paper help to interpret such empirical findings.
Appendix

Proof of proposition 1

The moments of \( \theta_m \) are key to the distribution of the predicted excess return, \( \hat{y}_{T+1} \), under different window sizes. To derive these, first note that

\[
\theta_m = \frac{\sum_{t=m}^{T_1} x_t^2}{\sum_{t=m}^{T_1} x_t^2 + \sum_{t=T_1+1}^T x_t^2 - \frac{d}{\chi^2_{\nu_i}(\lambda_1) + \chi^2_{\nu_2}(\lambda_2)},}
\]

where \( \chi^2_{\nu_i}(\lambda_i) \) is distributed as a non-central chi-squared with \( \nu_i \) degrees of freedom and the non-centrality parameter \( \lambda_i = \nu_i \mu_x^2 \). Recall that \( \nu_1 = T_1 - m + 1 \) and \( \nu_2 = T - T_1 \). Hence \( \theta_m \) has a doubly non-central beta distribution with parameters \( \nu_1/2 \) and \( \nu_2/2 \) and non-centrality parameters \( \lambda_1 \) and \( \lambda_2 \). Approximating each of the non-central \( \chi^2 \) variables in \( \theta_m \) and using Patnaik’s approximation (Patnaik (1949)), we have

\[
\theta_m \approx \left( \frac{\nu_{1/2} + 2 \lambda_1}{\nu_{1/2} + \lambda_1} \right) \left( \frac{\nu_{2/2} + \lambda_2}{\nu_{2/2} + 2 \lambda_2} \right) Beta(f_1, f_2),
\]

where

\[
f_i = \left( \frac{\nu_{i/2} + \lambda_i}{\nu_{i/2} + 2 \lambda_i} \right)^2 = \frac{\nu_i(1 + 2 \mu_x^2)}{2 + 8 \mu_x^2} \equiv \nu_i k;
\]

and \( k = (1 + 2 \mu_x^2)^2 / (2 + 8 \mu_x^2) \). Since \( \lambda_i = \nu_i \mu_x^2 \), this simplifies as follows

\[
\left( \frac{\nu_{1/2} + 2 \lambda_1}{\nu_{1/2} + \lambda_1} \right) \left( \frac{\nu_{2/2} + \lambda_2}{\nu_{2/2} + 2 \lambda_2} \right) = 1,
\]

so \( \theta_m \approx Beta(f_1, f_2) \) and \( E[\theta_m] \) and \( E[\theta^2_m] \) can be directly calculated from the moments of the (central) beta distribution:

\[
E[\theta_m] \approx \frac{f_1}{f_1 + f_2} = \frac{\nu_1}{\nu_1 + \nu_2} = \frac{\nu_1}{\nu},
\]

\[
E[\theta^2_m] \approx \frac{f_1(f_1 + 1)}{(f_1 + f_2)(f_1 + f_2 + 1)} = \frac{k \nu_1 (1 + k \nu_1)}{k(\nu_1 + \nu_2)(k(\nu_1 + \nu_2) + 1)} = \left( \frac{\nu_1}{\nu} \right) \left( 1 + k \nu_1 \right) \left( 1 + k \nu \right).
\]

\(^2\)See, for example, Johnson and Kotz (1970) pages 197-198.
Under the assumptions stated in Proposition 1, the conditional distribution of \( \hat{y}_{T+1} \) given the sequence of \( x \)'s \( \{x_1, \ldots, x_{T+1} \} \) is

\[
E[\hat{y}_{T+1}|x_1, x_2, \ldots, x_{T+1}] = \beta_2 x_{T+1} + (\beta_1 - \beta_2)\theta_m x_{T+1},
\]
or, unconditionally,

\[
E[\hat{y}_{T+1}] = \beta_2 \mu_x + (\beta_1 - \beta_2)E[\theta_m x_{T+1}].
\]

Since \( \theta_m \) only depends on \( x_1, \ldots, x_T \), it is independent of \( x_{T+1} \) and we have by the law of iterated expectations

\[
\begin{align*}
E[\hat{y}_{T+1}] &= \beta_2 \mu_x + (\beta_1 - \beta_2)E\{E[\theta_m x_{T+1}|x_1, \ldots, x_T]\}, \\
&= \beta_2 \mu_x + (\beta_1 - \beta_2)E\{\theta_m E[x_{T+1}|x_1, \ldots, x_T]\}, \\
&= \beta_2 \mu_x + (\beta_1 - \beta_2)\frac{\nu_1}{\nu}\mu_x, \\
&= \mu_x \left( \frac{\beta_1 \nu_1 + \beta_2 \nu_2}{\nu} \right).
\end{align*}
\]

Hence

\[
\begin{pmatrix}
y_{T+1} \\
\hat{y}_{T+1}
\end{pmatrix}
\sim II\mathcal{N}\left\{ \begin{pmatrix}
\beta_2 \mu_x \\
\beta_2 \mu_x + (\beta_1 - \beta_2) \left( \frac{\nu_1}{\nu} \right)
\end{pmatrix}, \Sigma \right\},
\]

where

\[
\Sigma = \begin{pmatrix}
\beta_2^2 \omega^2 + \sigma^2 & g \\
g & h^2
\end{pmatrix},
\]

and \( h^2 = V(\hat{y}_{T+1}) \) and \( g = Cov(y_{T+1}, \hat{y}_{T+1}) \). First consider the unconditional variance of \( \hat{y}_{T+1} \):

\[
h^2 = V(\hat{y}_{T+1}) = E[V(\hat{y}_{T+1}|x_1, \ldots, x_T, x_{T+1})] + V\left[ E(\hat{y}_{T+1}|x_1, \ldots, x_T, x_{T+1}) \right].
\]

Using the assumption that \( \sigma_1^2 = \sigma_2^2 = \sigma^2 \), the conditional variance of \( \hat{y}_{T+1} \) is given by

\[
V(\hat{y}_{T+1}|x_1, \ldots, x_T, x_{T+1}) = \frac{\sigma^2 x_{T+1}^2}{\sum_{t=m}^{T} x_t^2}.
\]

Therefore, using the expression for \( E[\hat{y}_{T+1}|x_1, x_2, \ldots, x_{T+1}] \) we have

\[
h^2 = \sigma^2 E\left( \frac{x_{T+1}^2}{\sum_{t=m}^{T} x_t^2} \right) + V(\beta_2 x_{T+1} + (\beta_1 - \beta_2)\theta_m x_{T+1}).
\]
The second term in this expression is given by

\[
V\{E(\hat{y}_{T+1}|x_1, \ldots, x_{T+1})\} = \beta_2^2 \omega^2 + (\beta_1 - \beta_2)^2 E[(\theta_m x_{T+1} - \mu_x \frac{\nu_1}{\nu})^2]
\]

\[
+ 2\beta_2 (\beta_1 - \beta_2) E[(x_{T+1} - \mu_x)(\theta_m x_{T+1} - \mu_x \frac{\nu_1}{\nu})]
\]

\[
= \beta_2^2 \omega^2 + (\beta_1 - \beta_2)^2 \left( \omega^2 \frac{1 + k\nu_1}{1 + k\nu} + \mu_x^2 \frac{\nu_2}{\nu(1 + k\nu)} \right)
\]

\[
+ 2\beta_2 (\beta_1 - \beta_2) \frac{\nu_1}{\nu} \omega^2.
\]

To evaluate \(E\left(\frac{x_{T+1}^2}{\sum_{t=m}^{T} x_t^2}\right)\), we first note that

\[
\frac{x_{T+1}^2}{\sum_{t=m}^{T} x_t^2} = \frac{(x_{T+1}/\omega)^2}{\nu \sum_{t=m}^{T} (x_t/\omega)^2/\nu} = \left(\frac{1}{\nu}\right) \left(\frac{(x_{T+1}/\omega)}{\chi^2_{\nu}(\lambda)/\nu}\right)^2
\]

and hence

\[
\frac{x_{T+1}^2}{\nu \sum_{t=m}^{T} x_t^2} = [t_{\nu}(\delta, \lambda)]^2,
\]

where \(t_{\nu}(\delta, \lambda) = (x_{T+1}/\omega)/(\chi^2_{\nu}(\lambda)/\sqrt{\nu})\) is distributed as a double non-central \(t\)-distribution with \(\nu\) degrees of freedom and the non-centrality parameters \(\delta = \mu_x/\omega\) and \(\lambda = \nu \mu_x^2\).

Using results in Johnson and Kotz (1970, p. 214, eq. 25) we now have

\[
E\left(\frac{x_{T+1}^2}{\sum_{t=m}^{T} x_t^2}\right) = \kappa_1(\lambda, \delta, \nu) = \left(\frac{1}{2}\right) \exp(-\frac{1}{2} \lambda)(1 + \delta^2) \times
\]

\[
\sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\lambda\right)^j \Gamma\left(\frac{1}{2}(\nu - 2) + j\right)}{j! \Gamma\left(\frac{1}{2} \nu + j\right)}.
\]

Using this expression the total variance of \(\hat{y}_{T+1}\) can be written as:

\[
h^2 = V(\hat{y}_{T+1}) = \sigma^2 \kappa_1(\lambda, \delta, \nu) + \beta_2^2 \omega^2 +
\]

\[
(\beta_1 - \beta_2)^2 \omega^2 \frac{(\nu_1)}{\nu} \left\{ \left( \frac{1 + k\nu_1}{1 + k\nu} \right) + \mu_x^2 \frac{\nu_2}{\omega^2 (1 + k\nu)} \right\}
\]

\[
+ 2\beta_2 (\beta_1 - \beta_2) \frac{\nu_1}{\nu} \omega^2.
\]

Also

\[
g \equiv Cov(y_{T+1}, \hat{y}_{T+1}) = E[(\beta_2 (x_{T+1} - \mu_x) + u_{T+1})(\beta_2 (x_{T+1} - \mu_x))]
\]

\[
+ (\beta_1 - \beta_2)(\theta_m x_{T+1} - \mu_x \frac{\nu_1}{\nu}) + x_{T+1} \frac{\sum_{t=m}^{T} x_t u_t}{\sum_{t=m}^{T} x_t^2}
\]

\[
= \beta_2^2 \omega^2 + \beta_2 (\beta_1 - \beta_2) \omega^2 (\frac{\nu_1}{\nu}).
\]
Noting that
\[
\Pr(y_{T+1} > 0) = \Phi \left( \frac{\beta_2 \mu_x}{\sqrt{\beta_2^2 \omega^2 + \sigma^2}} \right),
\]
\[
\Pr(\widetilde{y}_{T+1} > 0) = \Phi \left( \frac{\mu_x (\beta_1 \nu_1 + \beta_2 \nu_2)}{\nu h} \right),
\]
it is easy to compute the joint probability that \(y_{T+1} > 0\) and \(\widetilde{y}_{T+1} > 0\):

\[
A_H = \Pr(y_{T+1} > 0, \widetilde{y}_{T+1} > 0) = \int_{a_2 = -\mu_2}^{\infty} \int_{a_1 = -\mu_1}^{\infty} f(a_1, a_2) da_1 da_2,
\]
where
\[
\mu_1 = \mu_x \beta_2, \quad \mu_2 = \mu_x (\beta_1 \nu_1 + \beta_2 \nu_2) / \nu,
\]
\[
f(a_1, a_2) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} a^T \Sigma^{-1} a \right)
\]
and \(a = (a_1, a_2)'\). Similarly,

\[
A_F = \Pr(y_{T+1} < 0, \widetilde{y}_{T+1} > 0) = \int_{a_2 = -\mu_2}^{\infty} \int_{a_1 = -\infty}^{-\mu_1} f(a_1, a_2) da_1 da_2.
\]

Using the above results we have

\[
H - F = \frac{A_H}{\Phi(\mu_1 / \sqrt{\beta_2^2 \omega^2 + \sigma^2})} - \frac{A_F}{1 - \Phi(\mu_1 / \sqrt{\beta_2^2 \omega^2 + \sigma^2})}.
\]

**Bibliography**


Figure 1: Hit minus false alarm rate as a function of observation window