View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

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

M. Hashem Pesaran
Allan Timmermann

CESifo Working Paper No. 875
Category 10: Empirical and Theoretical Methods
February 2003

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the CESifo website: www.CESifo.de*

CESifo *Working Paper No. 875*

HOW COSTLY IS IT TO IGNORE BREAKS WHEN FORECASTING THE DIRECTION OF A TIME SERIES?

Abstract

Empirical evidence suggests that many 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 correlation between the forecast and the realization and 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 are likely to perform better over unconditional approaches that use expanding or rolling estimation windows provided that the break is reasonably large.

Keywords: sign prediction, estimation window, structural breaks.

JEL Code: C22, G10.

M. Hashem Pesaran
Trinity College
University of Cambridge
Sidgwick Avenue
Cambridge CB3 9DE
United Kingdom
mhp1@econ.cam.ac.uk

Allan Timmermann
University of California, San Diego
Department of Economics
9500 Gilman Drive
La Jolla, CA 92093-0508
U.S.A.

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. Three referees provided very helpful comments on the paper.

1. **Introduction**

Structural instability seems to characterize many forecasting models fitted to economic and financial data. In the most systematic study to date of a very large set of macroeconomic time series, Stock and Watson (1996) find evidence of structural instability in the majority of the series they consider. Many other studies have confirmed the presence of breaks in economic and financial time series; see, inter alia, Alogoskoufis and Smith (1991), Garcia and Perron (1996), Bai and Perron (199), Clements and Hendry (1998), Pesaran and Timmermann (2001) and Timmermann (2001).

Breaks or jumps in the parameters of forecasting models could arise from factors such as major changes in market sentiments, burst or creation of speculative bubbles, changes in monetary and debt management (for example, from money supply targeting to inflation targeting or from short-term to long-term debt instruments). The end of a national or global recession is another example that would be well suited to be modeled as a discrete shift in the parameters of the underlying data generating process.

Some models are even built around recurring shifts in the parameters of the underlying data generating process. For example, Hamilton (1989) studies Markov switching models that are driven by a latent variable process subject to discrete changes. Regime switches seem to characterize a host of financial time series including interest rates (Ang and Bekaert (2002), Driffill and Sola (1994) Gray (1996)) and stock market returns (Ang and Bekaert (2001), David and Veronesi (2001), Perez-Quiros and Timmermann (2000)). The parameters may be drawn either from a finite number of recurring states or from an expanding set of states as in the Bayesian change-point setup considered by Chib (1998).

In this paper we consider the problem of forecasting under breaks in the data generating process and characterize 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. We show this point more formally and quantify the factors that determine the gains from using a conditional

approach.

Structural instability in the data generating process could well have a significant impact on the performance of existing forecasting methods. Unfortunately, the implications for forecasting of such breaks have not been extensively explored although it is known that most econometric forecasting models perform very poorly around cyclical turning points. Turning points are of obvious interests to economists so it is important to gain a better understanding of how best to design and estimate forecasting models in the presence of parameter instability. In the presence of breaks, it is not clear, for example, that using the full data set to estimate a forecasting model leads to better predictions than if a carefully selected subset of recent data is used.

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 predictive accuracy 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.

Testing and estimation in the context of models that are subject to structural instability has been the subject of considerable research. Chow (1960) proposed an F-test for a single structural break in a linear regression model. This test assumes that the date of the break is known. Brown, Durbin and Evans (1975) developed Cusum and Cusum Squared tests to deal with the case where the time of the break is unknown. Recent work has extended these earlier tests in several directions to allow for multiple breaks, unit root dynamics and heteroskedasticity. Ploberger, Kramer and Kontrus (1989), Hansen (1992), Andrews (1993), Inclan and Tiao (1994), Andrews and Ploberger (1996), Chu, Stinchcombe and White (1996) and Bai and Perron (1998, 2002) develop tests for the consistent estimation of the size

and timing of breaks. Elliott and Mueller (2002) consider optimal tests for models with either many or relatively few breaks.

The plan of the paper is as follows. Section 2 introduces the break point model and derives analytical results for the mean and variance-covariance matrix of the joint distribution of the forecast and realization. Section 3 demonstrates how to quantify the market timing information in forecasting models that account for breaks relative to models that ignore these. Section 4 provides numerical examples that illustrate the analytical results. Section 5 concludes and discusses the empirical relevance of our findings. Mathematical proofs are provided in an Appendix.

2. **A Simple Break Model**

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

$$\begin{array}{rcll} y_t & = & \mathbf{x}_{t-1}'\boldsymbol{\beta}_1 + u_t, & u_t \sim IID(0,\sigma_1^2), \qquad t = 1,2,...,T_1 \\ & & \mathbf{x}_{t-1}'\boldsymbol{\beta}_2 + u_t, & u_t \sim IID(0,\sigma_2^2), \qquad t = T_1+1,...,T+1. \end{array} \tag{1}$$

y_t is the variable that is being predicted, $\mathbf{x}_{t-1}$ is the $p \times 1$ vector of pre-determined variables, $\boldsymbol{\beta}_i$ $(i = 1,2)$ are $p \times 1$ vectors of regression coefficients, and u_t is a serially uncorrelated error term that is independently distributed of $\mathbf{x}_s$ for all t and s, possibly with a shift in its variance from σ_1^2 to σ_2^2 at the time of the break point.

Suppose we are interested in predicting y_{T+1} conditional on period T information, $\Omega_T = (y_T, \mathbf{x}_T, y_{T-1}, \mathbf{x}_{T-1}, ...)$, and the knowledge of the break at point $t = T_1$, but not the size of the break. The question is whether we should use only post break observations or is it optimal to use some pre-break observations as well? Denote the pre- and post-break observations by $v_1 \leq T_1$ and $v_2 = T - T_1$, respectively, and denote the Ordinary Least Squares (OLS) estimate of the regression coefficients estimated over the sample $v = v_1 + v_2$ by $\widehat{\boldsymbol{\beta}}_m$. Then the point forecast of y_{T+1} is given by[1]

$$\widehat{y}_{T+1} = \mathbf{x}_T'\widehat{\boldsymbol{\beta}}_m,$$

[1]An alternative predictor of the sign of y_{T+1} is developed by Skouras 1999) who considers estimating $\boldsymbol{\beta}$ by maximizing

$$\sum_{t=T_1-v_1+1}^{T} I\left[y_t\left(\mathbf{x}_{t-1}'\boldsymbol{\beta}\right)\right]$$

where $I[A]$ is an indicator function that takes the value of unity if $A > 0$ and zero otherwise.

where

$$\widehat{\boldsymbol{\beta}}_m = \left(\sum_{t=m}^{T} \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \left(\sum_{t=m}^{T} \mathbf{x}_{t-1} \mathbf{y}_t \right),$$

and $m = T_1 - v_1 + 1 \geq 1$.

In general it is complicated to derive analytical expressions for measures of predictive accuracy. However, in the very simple case where

$$\begin{aligned} p &= 1, \sigma_1^2 = \sigma_2^2 = \sigma^2, \\ u_t &\sim IIN(0, \sigma^2), \, x_t \sim IIN(\mu_x, \omega^2) \end{aligned}$$

we can derive tractable results for how the correlation between y_{T+1} and $\widehat{y}_{T+1}$ depends on the size of the post-break window (v_2) relative to the size of the pre-break window (v_1).

In this simple case

$$y_{T+1} = \beta_2 x_T + u_{T+1}, \tag{2}$$

$$\widehat{y}_{T+1} = \beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T + x_T \left(\frac{\sum_{t=m}^{T} x_{t-1} u_t}{\sum_{t=m}^{T} x_{t-1}^2} \right), \tag{3}$$

where

$$\theta_m = \theta_m(T_1, T) = \frac{\sum_{t=m}^{T_1} x_{t-1}^2}{\sum_{t=m}^{T} x_{t-1}^2}.$$

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,

$$\begin{pmatrix} u_t \\ x_{t-1} \end{pmatrix} \sim IIN \left(\begin{pmatrix} 0 \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \omega^2 \end{pmatrix} \right).$$

Then the mean and variance-covariance matrix of ($y_{T+1}, \widehat{y}_{T+1}$) are given by

$$E(y_{T+1}) = \mu_1 = \beta_2 \mu_x, \tag{4}$$

$$E(\widehat{y}_{T+1}) = \mu_2 = \beta_2 \mu_x + \left(\frac{v_1}{v} \right) (\beta_1 - \beta_2)\mu_x, \tag{5}$$

$$Var \begin{pmatrix} y_{T+1} \\ \widehat{y}_{T+1} \end{pmatrix} = \boldsymbol{\Sigma} = \begin{pmatrix} \beta_2^2 \omega^2 + \sigma^2 & g \\ g & h^2 \end{pmatrix}, \tag{6}$$

where

$$h^2 = V(\widehat{y}_{T+1}) = \sigma^2\phi + \beta_2^2\omega^2 + \omega^2(\frac{\nu_1}{\nu})(\beta_1 - \beta_2)^2\psi + 2\beta_2(\beta_1 - \beta_2)\omega^2(\frac{\nu_1}{\nu}). \quad (7)$$

and

$$g \equiv Cov(y_{T+1}, \widehat{y}_{T+1}) = \beta_2^2\omega^2 + (\frac{\nu_1}{\nu})\beta_2(\beta_1 - \beta_2)\omega^2. \quad (8)$$

The results in Proposition 1 are exact. The parameters ϕ and ψ are complicated functions of the window size parameters and the means and variances of the underlying series. They are defined by (20) and (19) in the Appendix.

The proposition shows how the correlation between y_{T+1} and $\widehat{y}_{T+1}$ depends on the size of the pre-break (v_1) and post-break (v_2) window and some intuition can already be gathered from these expressions. Suppose that a window of $v_1 \geq 1$ pre-break observations has been used to estimate the regression coefficient of x. Then the estimated mean of y_{T+1}, namely $\mu_2 = \beta_2\mu_x + \left(\frac{v_1}{\nu}\right)(\beta_1 - \beta_2)\mu_x$, will be biased, although the forecast error variance will also be smaller than if only post-break observations were used ($v_1 = 0$). Likewise, the larger is the break size, $|\beta_1 - \beta_2|$, and the smaller the post-break relative to pre-break observations (v_2/v_1) the weaker is the correlation between predicted and realized values and thus the lower the sign test statistic.[2]

3. **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.

The probability of correctly predicting the sign of y_{T+1} is given by $\Pr(y_{T+1}\widehat{y}_{T+1} > 0)$. 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

[2]Note that $v_1/v = 1/(1 + v_2/v_1)$.

convenient to work with. Granger and Pesaran (2000) show that this market timing statistic can also be written as

$$PT = \frac{\sqrt{n}(H - F)}{\left(\frac{\hat{\pi}(1-\hat{\pi})}{\pi(1-\pi)}\right)^{1/2}}, \tag{9}$$

where n is the number of observations in the forecasting sample, H is the "hit rate" and F is the "false alarm rate", which are defined as

$$\begin{aligned} H &= \frac{\Pr(y_{T+1} > 0,\ \widehat{y}_{T+1} > 0)}{\Pr(y_{T+1} > 0)}, \\ F &= \frac{\Pr(y_{T+1} < 0,\ \widehat{y}_{T+1} > 0)}{\Pr(y_{T+1} < 0)}. \end{aligned} \tag{10}$$

Finally $\pi = \Pr(y_{T+1} > 0)$, and $\hat{\pi} = \Pr(\widehat{y}_{T+1} > 0)$ 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.

The unconditional distribution of $\widehat{y}_{T+1}$ is a mixture normal and can be derived in the following manner. Conditional on $\mathbf{X}_{mT} = (x_{m-1}, x_m, ..., x_T)$ we have

$$\widehat{y}_{T+1} \mid \mathbf{X}_{mT} \sim N(\mu_{mT}, \sigma^2_{mT}),$$

where

$$\begin{aligned} \mu_{mT} &= \beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T, \\ \sigma^2_{mT} &= x^2_T \frac{\sigma^2}{\sum_{t=m}^{T} x^2_{t-1}}. \end{aligned}$$

Denote the joint distribution of $\mathbf{X}_{mT}$ by $f(\mathbf{X}_{mT})$. Then the unconditional distribution of $\widehat{y}_{T+1}$ is given by

$$f(\widehat{y}_{T+1}) = \int_{\mathbf{X}_{mT}} \varphi\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right) f(\mathbf{X}_{mT}) d\mathbf{X}_{mT},$$

where

$$\varphi\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right) = \left(2\pi\sigma^2_{mT}\right)^{-1/2} \exp\left[\frac{-1}{2}\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right)^2\right].$$

The probability $\Pr(\widehat{y}_{T+1} > 0)$ can now be computed using $f(\widehat{y}_{T+1})$:

$$\Pr(\widehat{y}_{T+1} > 0) = \int_{y=0}^{\infty} \int_{\mathbf{X}_{mT}} \varphi\left(\frac{y - \mu_{mT}}{\sigma_{mT}}\right) f(\mathbf{X}_{mT}) d\mathbf{X}_{mT} dy.$$

In the case where $x_t \sim IIN(\mu_x, \omega^2)$ we have

$$f(\mathbf{X}_{mT}) = \prod_{t=m}^{T} \varphi\left(\frac{x_{t-1} - \mu_x}{\omega}\right).$$

Computation of $\Pr(\widehat{y}_{T+1} > 0)$ in general involves high dimensional multiple integrals. But under $x_t \sim IIN(\mu_x, \omega^2)$ the integral for $f(\widehat{y}_{T+1})$ can be written in terms of x_T , $\sum_{t=m}^{T_1} x_{t-1}^2/\omega^2$ and $\sum_{t=T_1+1}^{T} x_{t-1}^2/\omega^2$ which are independently distributed with known distributions, namely $IIN(\mu_x, \omega^2)$, $\chi^2_{v_1}(\lambda_1)$ and $\chi^2_{v_2}(\lambda_2)$, respectively, where $\chi^2_{v_i}(\lambda_i)$ denotes a non-central chi-squared distribution with the non-centrally parameter, $\lambda_i = v_i (\mu_x/\omega)^2$. In the appendix we show the following result:

Proposition 2

Under the assumptions of Proposition 1, the probabilitities underlying the sign prediction problem are given as follows:

$$\begin{aligned}
\Pr(\widehat{y}_{T+1} > 0) &= \int_{y=0}^{\infty} \int_{z=-\infty}^{\infty} \int_{\chi^2_{v_1}=0}^{\infty} \int_{\chi^2_{v_2}=0}^{\infty} \varphi\left(\frac{y - \mu_{mT}}{\sigma_{mT}}\right) \\
&\quad \times \varphi(z)\, f(\chi^2_{v_1}(\lambda_1)) f(\chi^2_{v_2}(\lambda_2)) dz\; d\chi^2_{v_1} d\chi^2_{v_2} dy. \\
\Pr(y_{T+1} &> 0) = \varphi\left(\frac{\beta_2 \mu_x}{\sqrt{\beta_2^2 \omega^2 + \sigma^2}}\right) \\
\Pr(y_{T+1} &> 0, \widehat{y}_{T+1} > 0) = \int_{y=0}^{\infty} \int_{\widehat{y}=0}^{\infty} f(y, \widehat{y}) dy d\hat{y}.
\end{aligned}$$

where

$$\begin{aligned}
f(y_{T+1}, \widehat{y}_{T+1}) &= \int_{z=-\infty}^{\infty} \int_{\chi^2_{v_1}=0}^{\infty} \int_{\chi^2_{v_2}=0}^{\infty} \varphi\left(\frac{y_{T+1} - \beta_2(\omega\; z + \mu_x)}{\sigma}\right) \\
&\quad \times \varphi\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right) \varphi(z)\, f(\chi^2_{v_1}(\lambda_1)) f(\chi^2_{v_2}(\lambda_2)) dz\; d\chi^2_{v_1} d\chi^2_{v_2}.
\end{aligned}$$

This proposition characterizes the market timing value of the forecast as a function of the parameters of the underlying data generating process and the design

parameters of the estimation window. The significance of the proposition lies in its reduction of a high-dimensional problem to a far more tractable problem of a much lower dimension. In general, computation of the sign probabilities require integration over an $(T-m+1)$ dimensional space which, by the proposition, can be reduced to integration over a much smaller three or four dimensional space.

In practice the computations of the sign probabilities is best be carried out by stochastic simulations. To compute $\Pr(\widehat{y}_{T+1}>0)$ numerically, generate $\mathbf{X}_{mT}^{(i)}$, $\mathbf{U}_{mT}^{(i)}$ as draws from the assumed distributions of x_t and u_t, where $\mathbf{U}_{mT}=(u_m,u_{m+1},....,u_{T+1})$ and as before $\mathbf{X}_{mT}=(x_{m-1},x_m,....,x_T)$. Then use these draws to compute $\widehat{y}_{T+1}^{(i)}$ and finally approximate $\Pr(\widehat{y}_{T+1}>0)$ by

$$\Pr{}_R(\widehat{y}_{T+1}>0)=\frac{1}{R}\sum_{i=1}^{R}I\left(\widehat{y}_{T+1}^{(i)}\right),$$

where $I(a)$ takes the value of unity if $a>0$ and 0 otherwise. To compute the joint probabilities, for given draws $\mathbf{X}_{mT}^{(i)}$, $\mathbf{U}_{mT}^{(i)}$, compute $\widehat{y}_{T+1}^{(i)}$ and $y_{T+1}^{(i)}$ as

$$y_{T+1}^{(i)} = \beta_2 x_T^{(i)}+u_{T+1}^{(i)}, \tag{11}$$

$$\widehat{y}_{T+1}^{(i)} = \beta_2 x_T^{(i)}+(\beta_1-\beta_2)\theta_m^{(i)}x_T^{(i)}+x_T^{(i)}\left(\frac{\sum_{t=m}^{T}x_{t-1}^{(i)}u_t^{(i)}}{\sum_{t=m}^{T}\left(x_{t-1}^{(i)}\right)^2}\right), \tag{12}$$

where

$$\theta_m^{(i)}=\frac{\sum_{t=m}^{T_1}\left(x_{t-1}^{(i)}\right)^2}{\sum_{t=m}^{T}\left(x_{t-1}^{(i)}\right)^2}.$$

Then

$$\Pr{}_R(y_{T+1}>0,\widehat{y}_{T+1}>0)=\frac{1}{R}\sum_{i=1}^{R}I\left(\widehat{y}_{T+1}^{(i)}\right)I\left(y_{T+1}^{(i)}\right),$$

$$\Pr{}_R(y_{T+1}<0,\widehat{y}_{T+1}>0)=\frac{1}{R}\sum_{i=1}^{R}I\left(\widehat{y}_{T+1}^{(i)}\right)I\left(-y_{T+1}^{(i)}\right).$$

These probabilities can easily be computed under a break in error variances ($\sigma_1\neq\sigma_2$) or when the regressors are serially correlated, or the errors are t-distributed and subject to conditional volatility.

The next section provides some numerical illustrations of how parameter breaks affect market timing under three common types of estimation window, namely an expanding, a rolling and a post-break window.

4. **Numerical Results**

The propositions in the previous sections allow us to determine the exact value of the correlation between the forecast and the realization as well as the 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. In this section we compare the predictive accuracy of three different methods for window size determination, namely a post-break method that only uses observations after the break, a fixed-length rolling window method and an expanding window method.

The results presented in this section assume that $\beta_2 = 1$, $\mu_x = 0.5$, $\sigma = 6$ and $\omega = 1.5$. 100 pre-break observations are assumed to be available. 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 $\nu = 60$ 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).

4.1. *Correlations between Forecasts and Realizations*

Figures 1-3 use Proposition 1 to plot the correlation between y_{T+1} and $\widehat{y}_{T+1}$ as a function of the value of β_1 which tracks the break size $|\beta_1 - \beta_2|$. Since β_2 is kept constant and the break size is varied by changing the value of the pre-break parameter, β_1, the correlation is independent of break size for the post-break window method. In contrast, the expanding and rolling window methods mix the post- and pre-break data samples so the correlations arising from these approaches deteriorate as the difference between β_1 and β_2 increases.

First consider the case with 10 post-break observations (Figure 1). In the absence of a break, the expanding window method is most efficient and thus produces the highest correlations for zero or very small breaks. For these situations the correlation between y_{T+1} and $\hat{y}_{T+1}$ exceeds 0.2. As the break size grows larger than one, the post-break method starts to dominate since it is not affected by the resulting bias due to using pre-break data points. The rolling window method also does

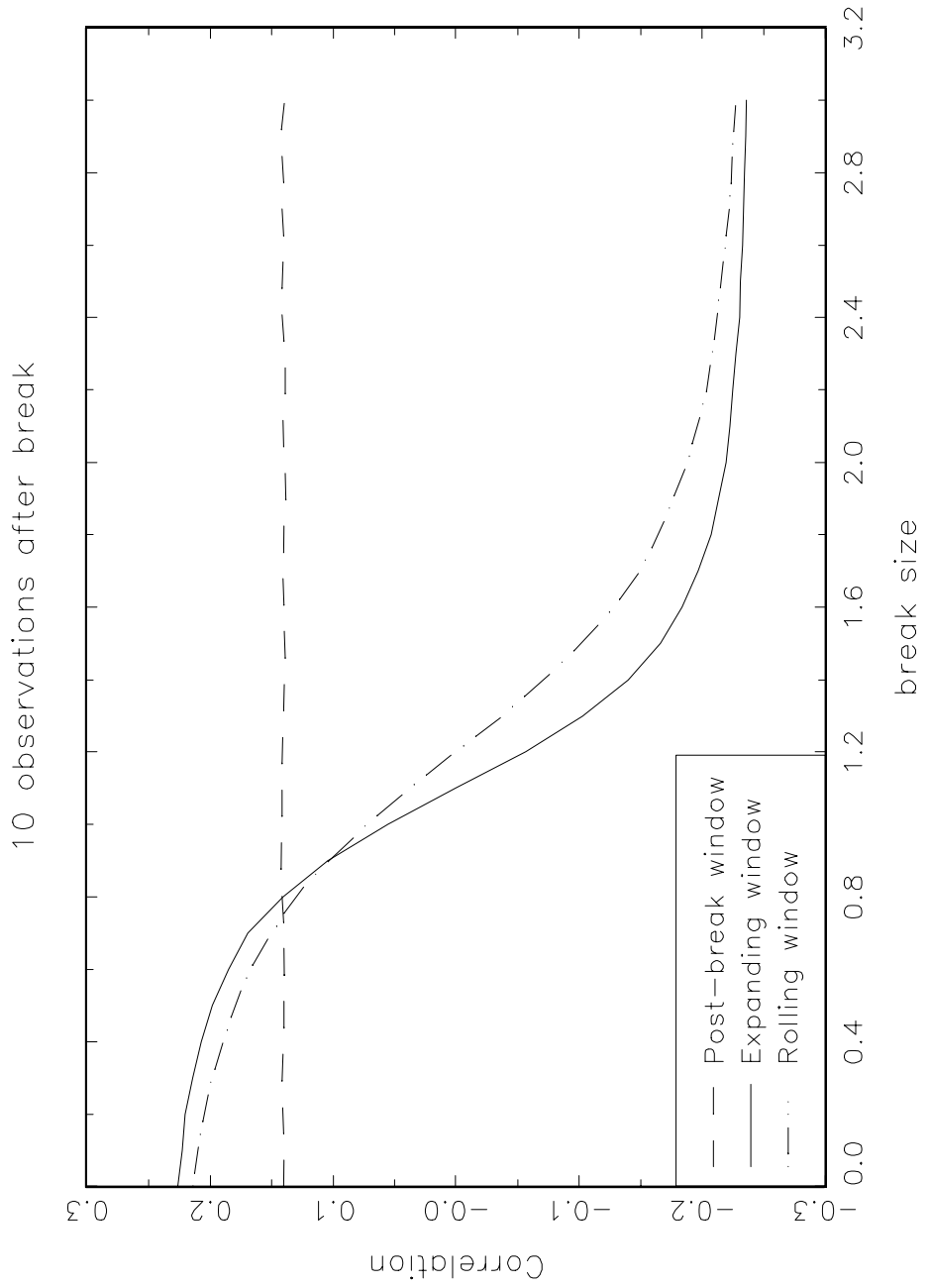

Figure 1: Correlation between realization, forecast $(y_{T+1}, \hat{y}_{T+1})$

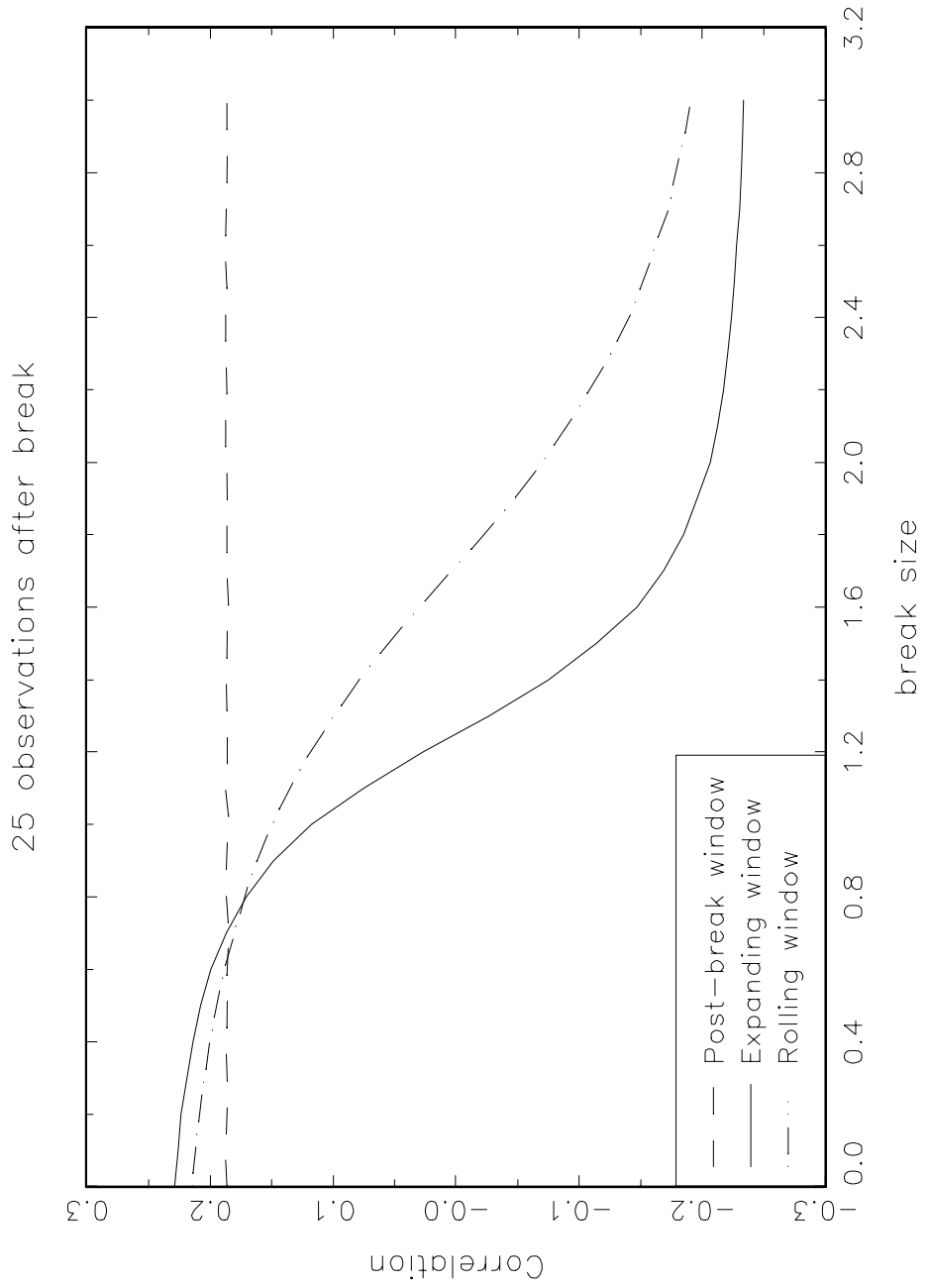

Figure 2: Correlation between realization, forecast $(y_{T+1}, \hat{y}_{T+1})$

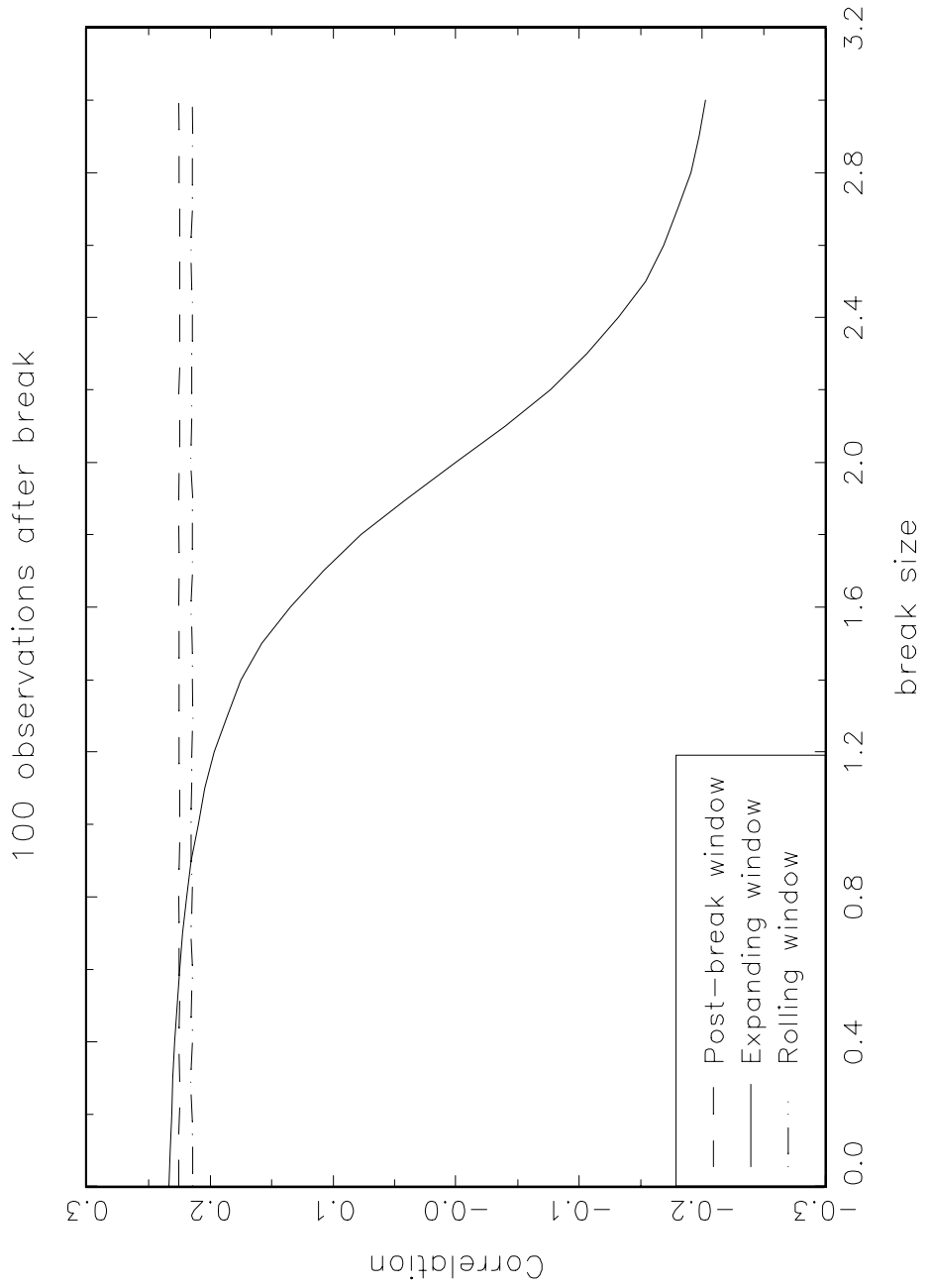

Figure 3: Correlation between realization, forecast $(y_{T+1}, \hat{y}_{T+1})$

better than the expanding window since it is less biased as it relies on relatively fewer pre-break data points.

As the distance to the most recent break grows to 25 observations (Figure 2) or 100 observations (Figure 3), the post-break method increasingly starts to dominate the other approaches and it is only for very small breaks that there is a gain from using an expanding window. Even when a gain is possible from using a rolling or an expandind window, the gain tends to be very marginal. Notice that the post-break window method always dominates the rolling window in Figure 3 since the latter is inefficient as it does not use all post-break data.

4.2. *Market Timing Values*

Figures 4-6 use proposition 2 to show how the market timing information in the predicted variable depends on whether an expanding or rolling window or a conditional break point method is used. The figures plot the hit minus false alarm rate $(H - F)$ as a function of the break size. We compute the values of this market timing statistic using the stochastic simulation methods described in the previous section, setting the number of replications, R, equal to one million. This is sufficiently large to make the results quite precise.

Figure 4 again assumes a small post break window of 10 observations, namely $v_2 = 10$. Again β_2 is kept constant and the break size is varied by changing the value of the pre-break parameter, β_1. This means that the hit minus false alarm rate is independent of break size for the post-break window method while it deteriorates for the expanding and rolling window methods that mix the post- and pre-break data samples.

In the absence of a break ($\beta_1 = \beta_2$), as to be expected the expanding and rolling windows produce better sign predictions than the post-break window which wrongly ignores pre-break data. However, as the break size increases, the post-break 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. In fact these methods generate a negative sign statistic for break sizes above one.

Figure 5 considers a longer post-break window and sets $v_2 = 25$. If only post-break information is used, $H - F$ is constant around 12 percent irrespective of break size. This is approximately 4% higher than in Figure 1 where less data was

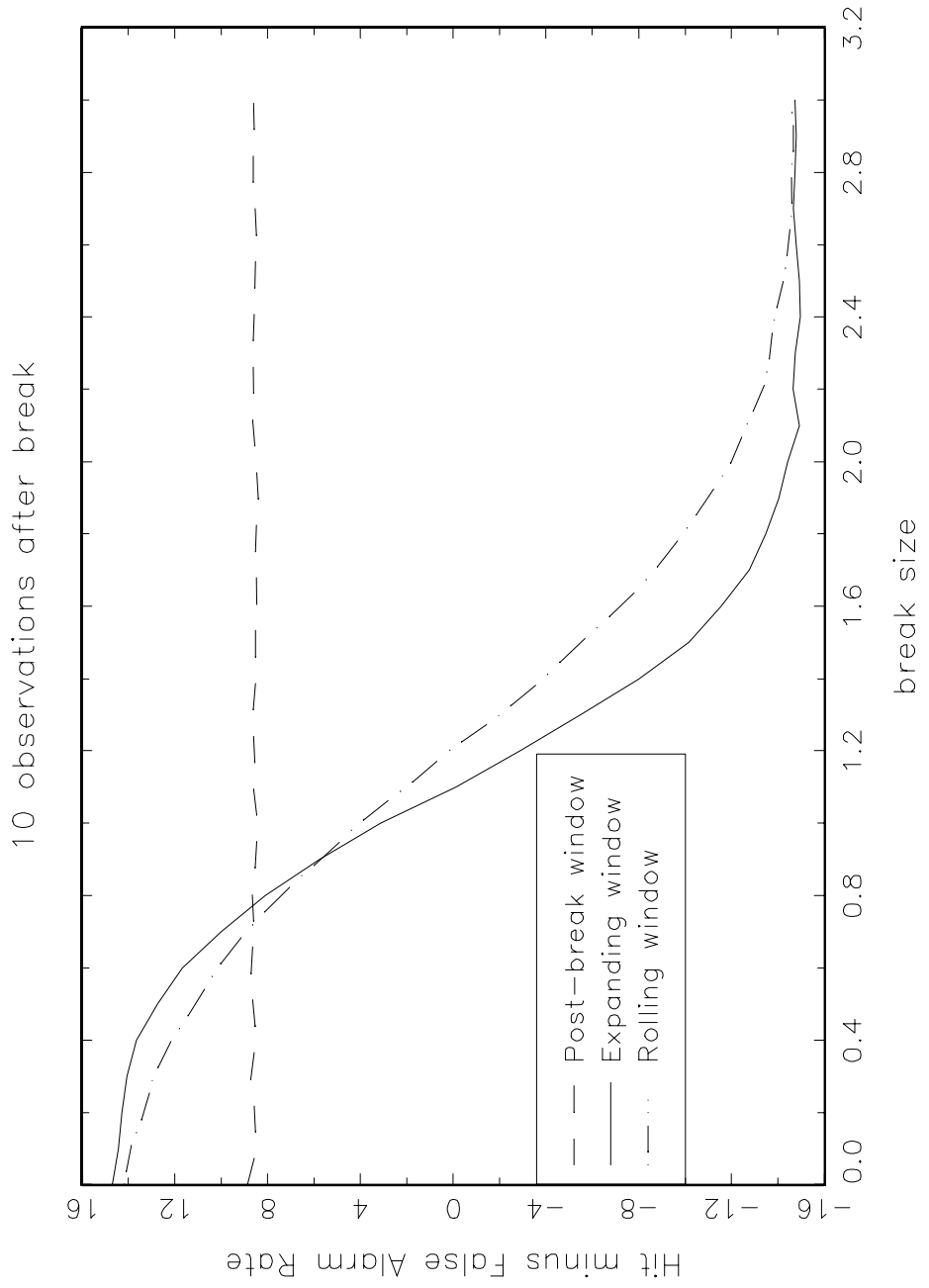

Figure 4: Hit minus false alarm rate (identical pre- and post-break variance)

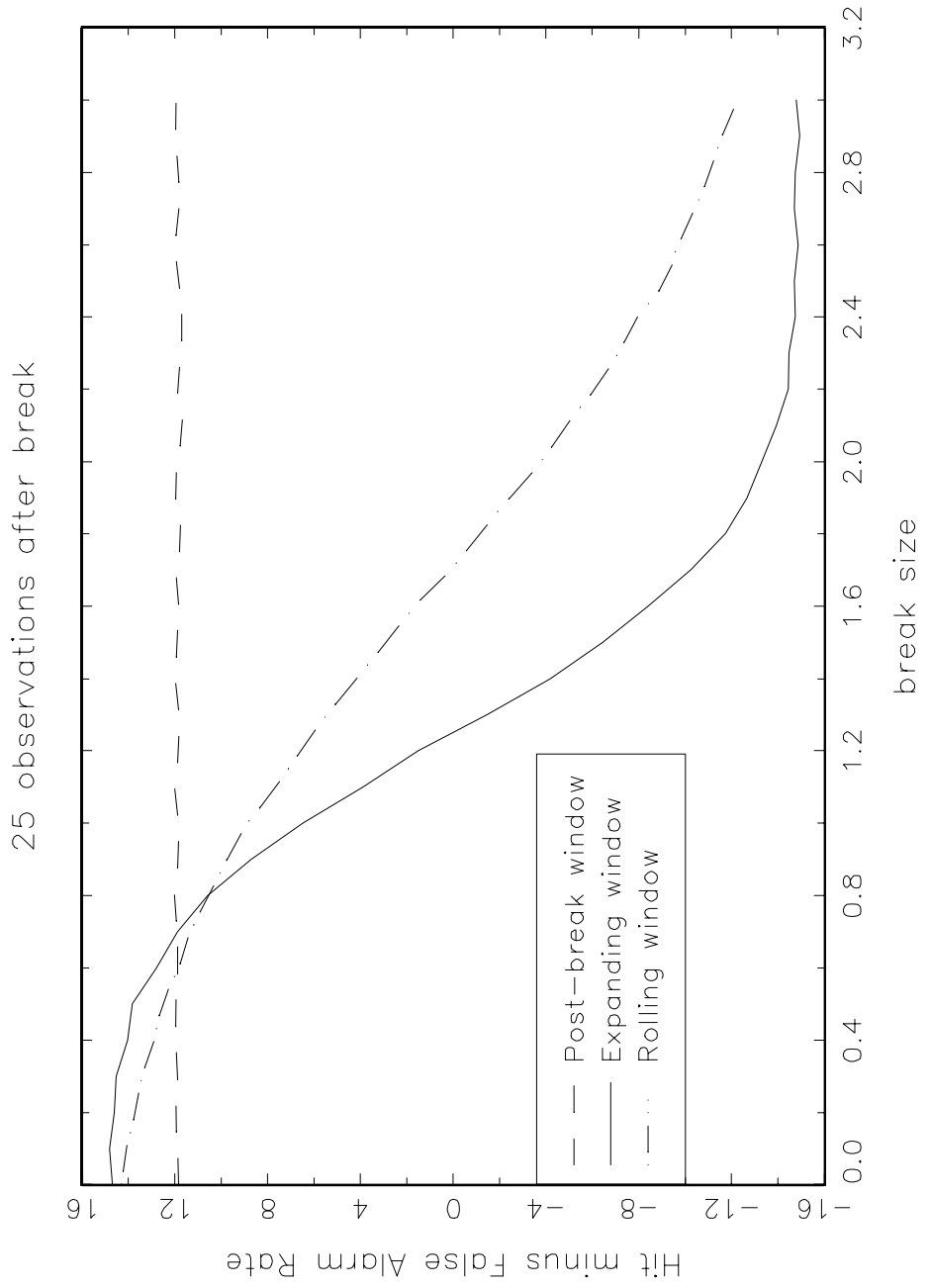

Figure 5: Hit minus false alarm rate (identical pre- and post-break variance)

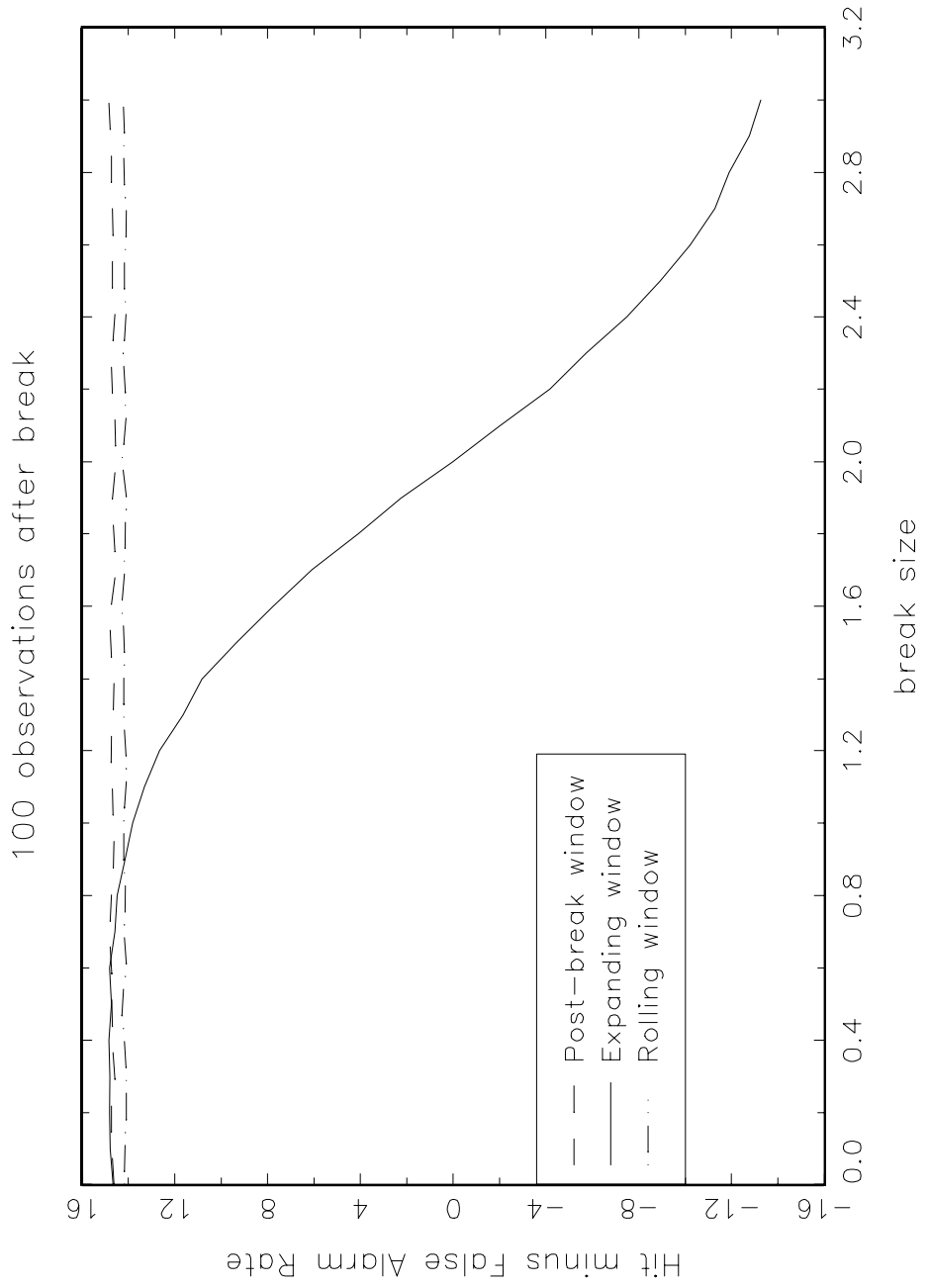

Figure 6: Hit minus false alarm rate (identical pre- and post-break variance)

available for parameter estimation ($v_2 = 10$). The expanding window generates negative values of the market timing statistic for a break size above 1.2. The rolling window only does marginally better. As the break size $|\beta_1 - \beta_2|$ increases, the hit minus false alarm rate continues to decline for the expanding and rolling windows, suggesting that a forecasting model that fails to account for a break can lead to a severe deterioration in market timing performance.

Figure 6 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 hit minus false alarm rate rises to 14 percent for the post-break method. Although the rolling window does not use any pre-break data,[3] it still underperforms slightly relative to the post-break method since it does not use all 100 post-break data points and hence is inefficient. However, for break sizes 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 relies on all pre-break data points quickly deteriorates as the bias effect takes over.

So far, our results have assumed that there is no break in the variance. In some situations, we would expect that $\sigma_2^2 \neq \sigma_1^2$. As noted earlier, allowing for a break in the variance is feasible provided that we resort to numerical methods for evaluating the variance of the forecast, $\widehat{y}_{T+1}$.[4] To study the effect of a simultaneous break in the mean and variance of the forecasting model, we consider two scenarios. In both cases, the pre-break volatility, σ_1, is kept fixed at 6. Under the first scenario, we lower the volatility of the forecast error, σ_2 from 6 to 3, thus increasing the predictive R^2 and bringing down the noise in the forecasting equation. Under the second scenario, we increase σ_2 to 9, thus bringing down the predictive R^2.

In figures 7-9 we assume $\sigma_1 = 6, \sigma_2 = 3$. First consider Figure 7. Despite having only 10 post-break data points, the post-break window method generates a hit minus false alarm rate of 24%. This is higher than the market timing values produced by the rolling and expanding window methods provided that the break size is greater than 0.6. For breaks of this size or larger, the forecasts associated with the unconditional window determination methods are heavily biased. With 25 or 100 post-break data points (Figures 8-9), $H - F$ rises to a level around 30% and the post-break method now dominates for all but the very smallest break sizes.

[3]The window size (60) is less than the time since the break (100).

[4]See the last Section of the Appendix.

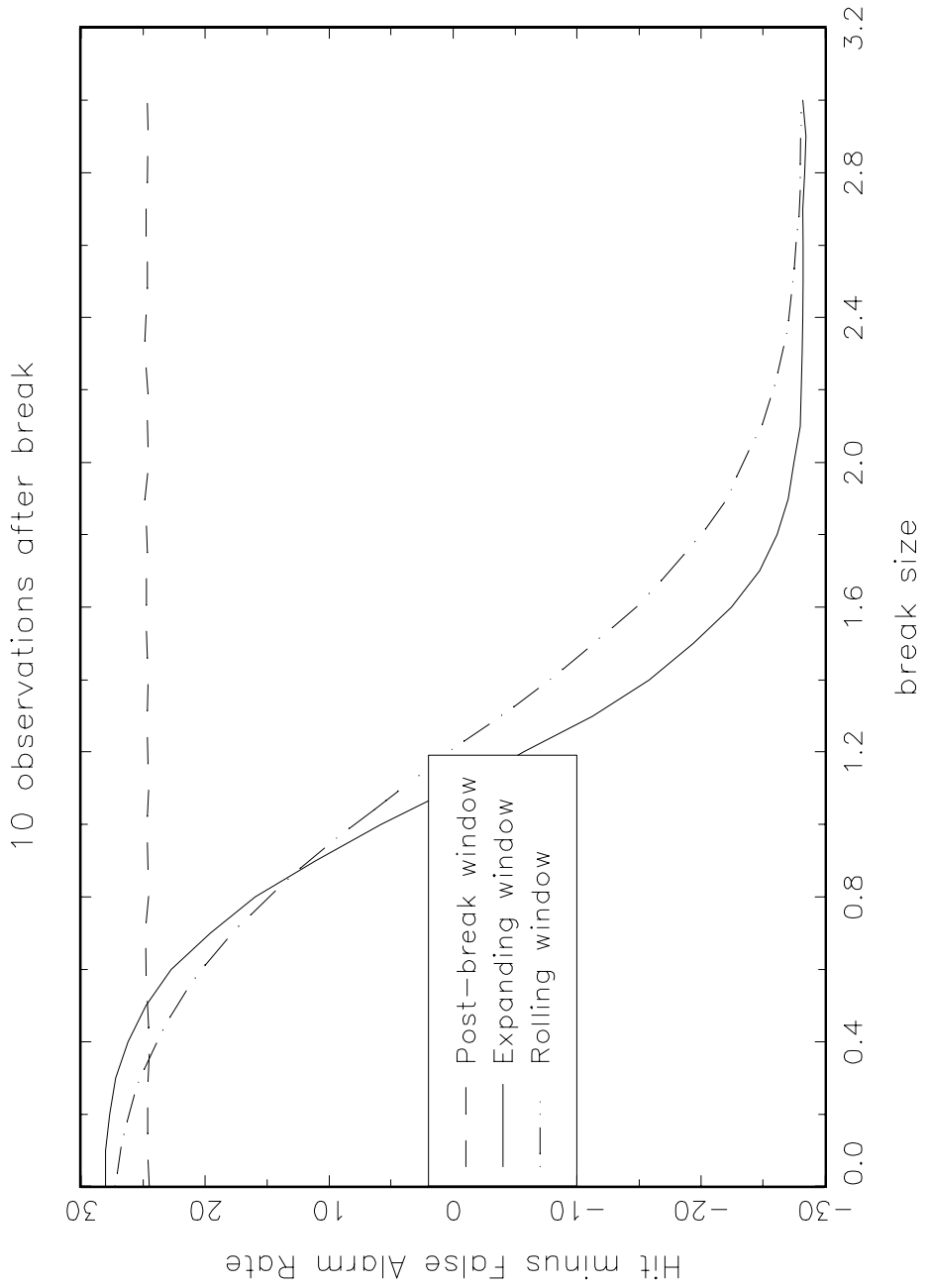

Figure 7: Hit minus false alarm rate (lower post-break variance)

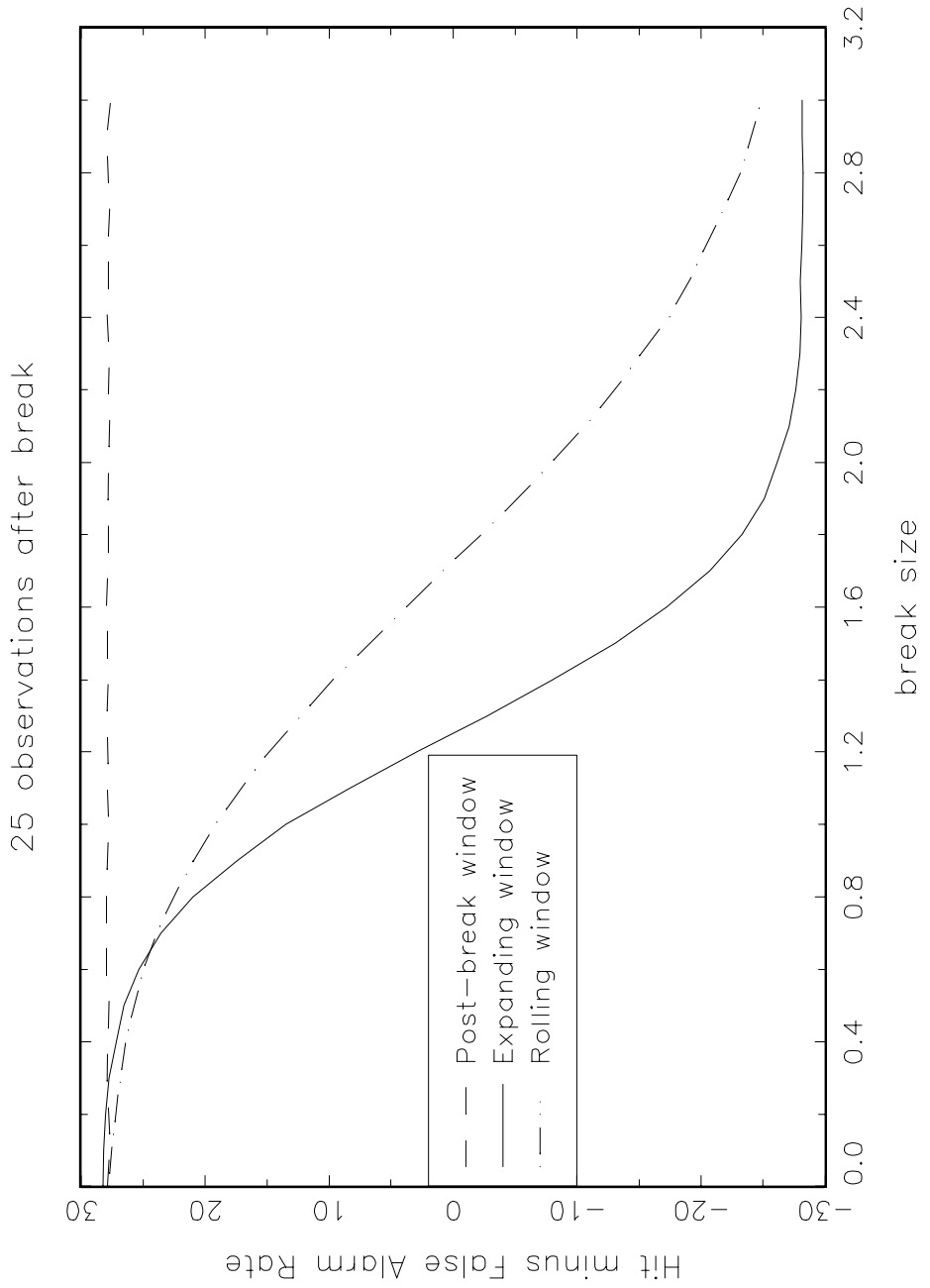

Figure 8: Hit minus false alarm rate (lower post-break variance)

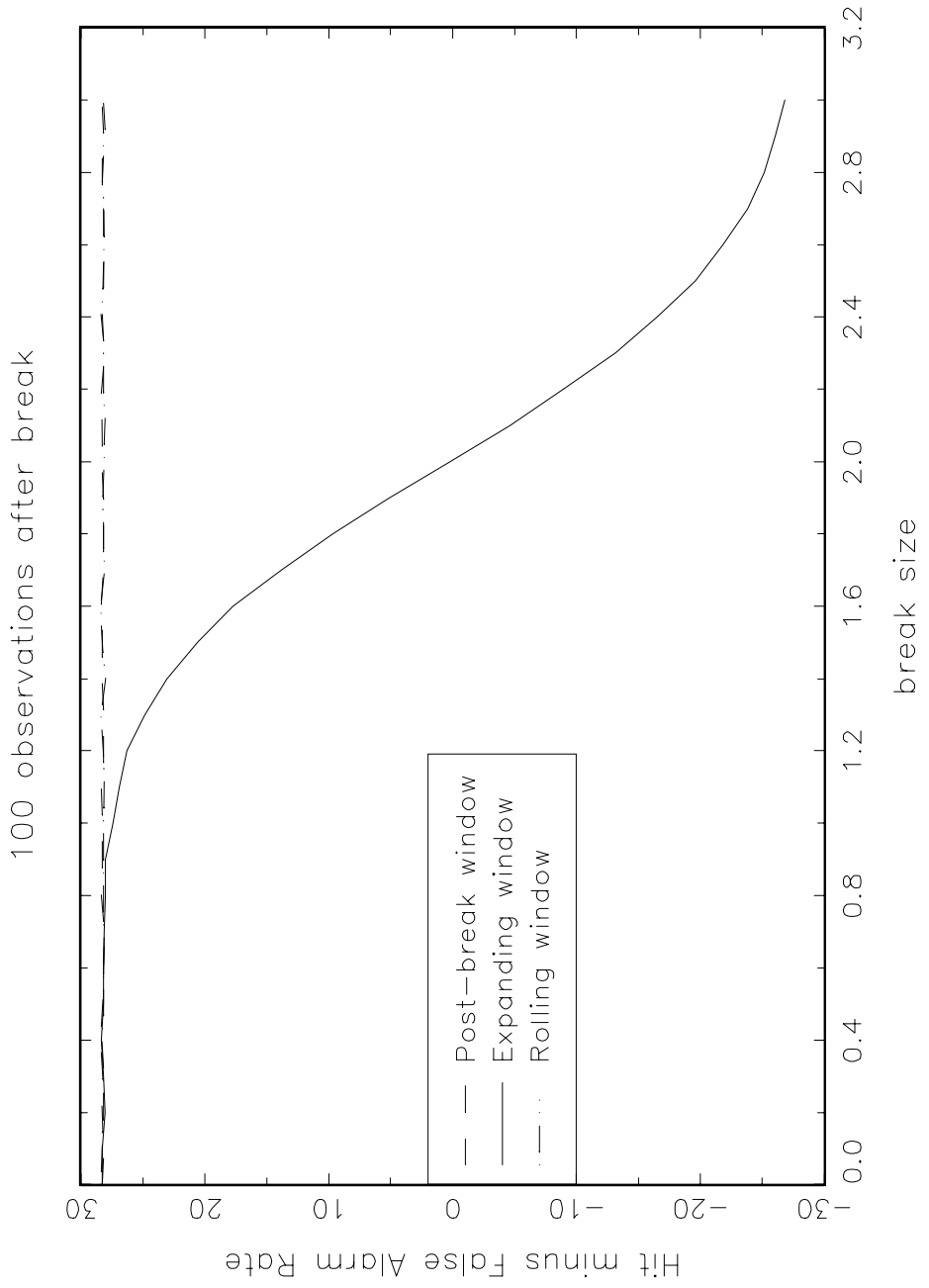

Figure 9: Hit minus false alarm rate (lower post-break variance)

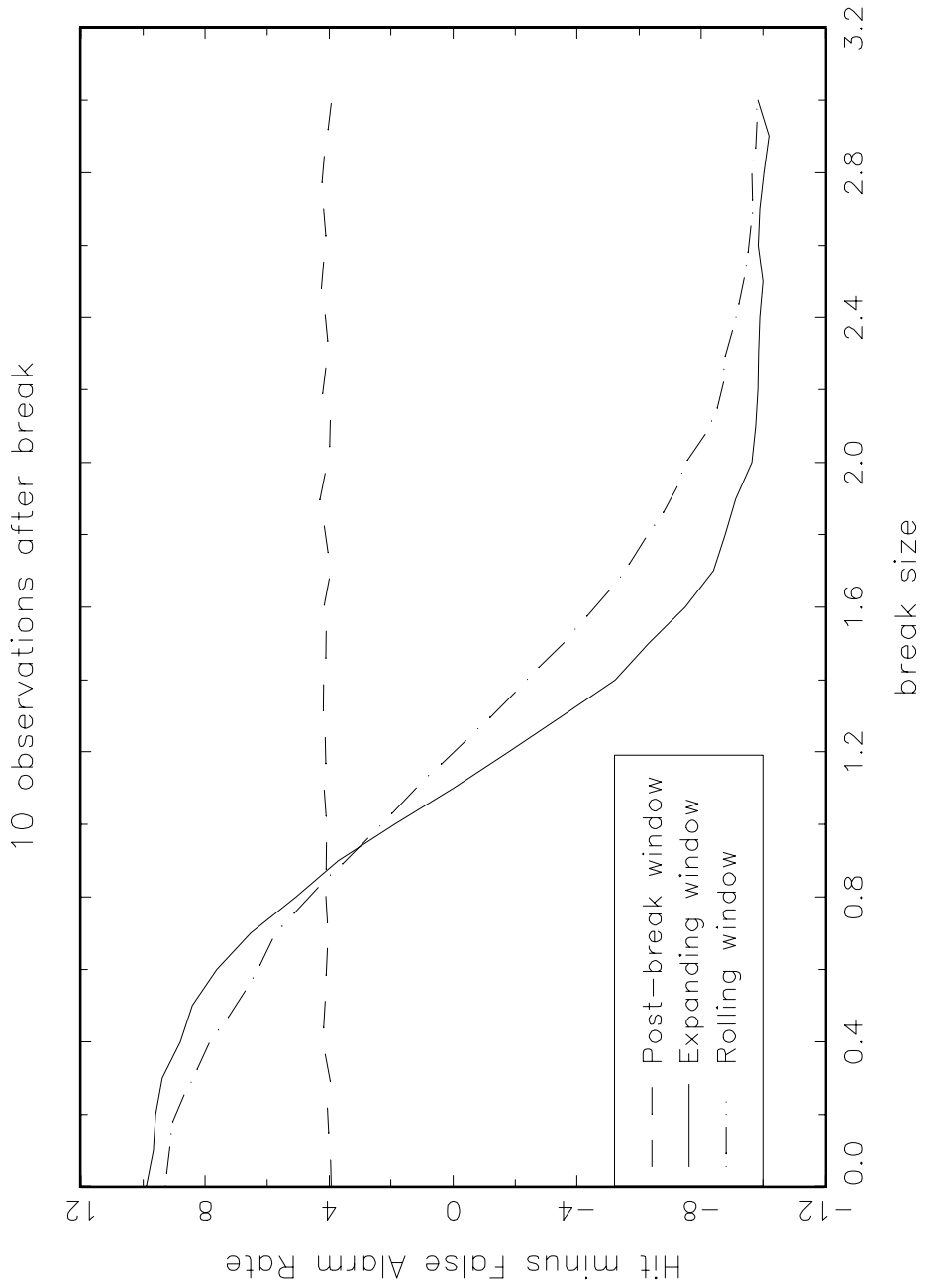

Figure 10: Hit minus false alarm rate (higher post-break variance)

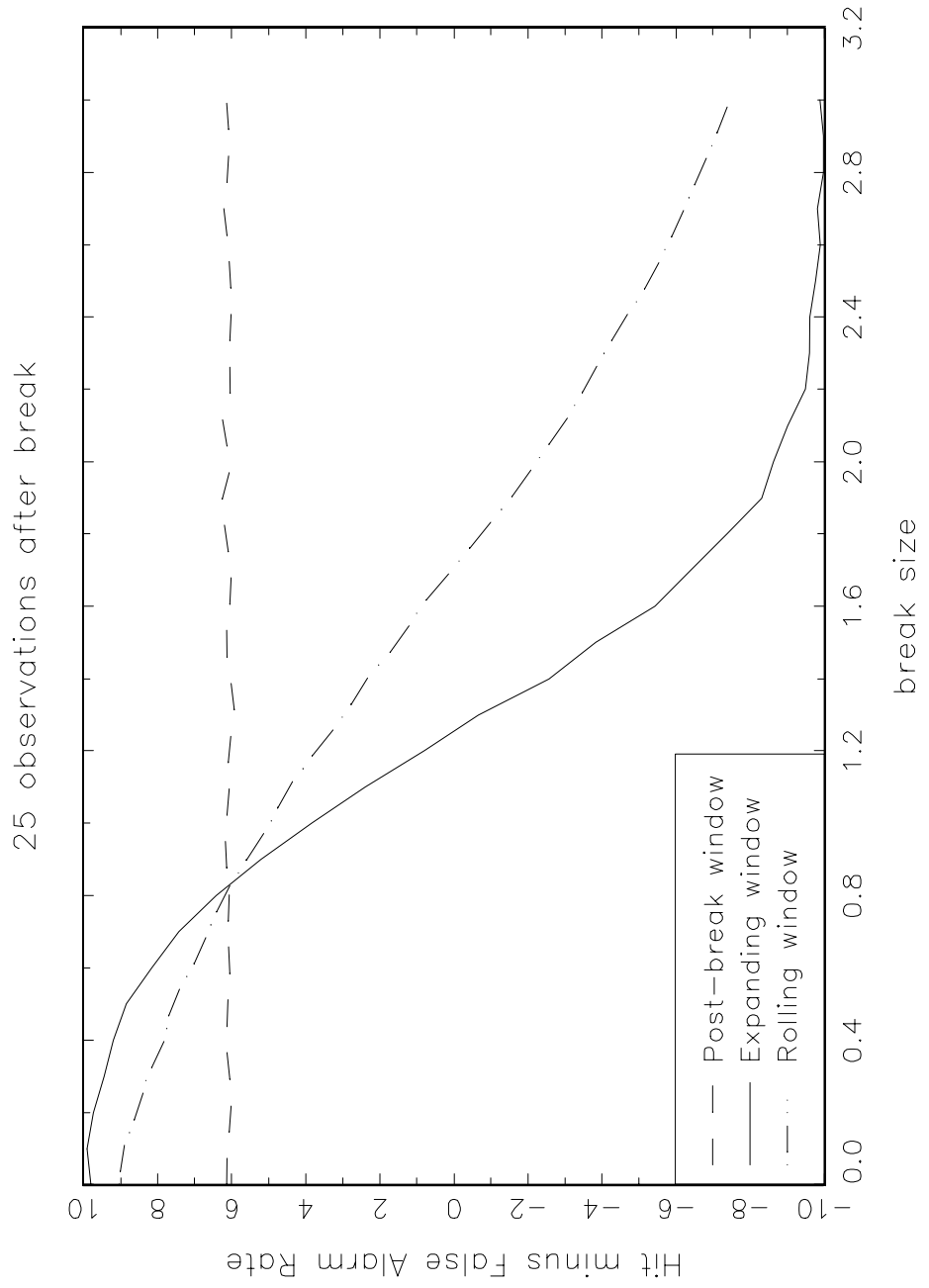

Figure 11: Hit minus false alarm rate (higher post-break variance)

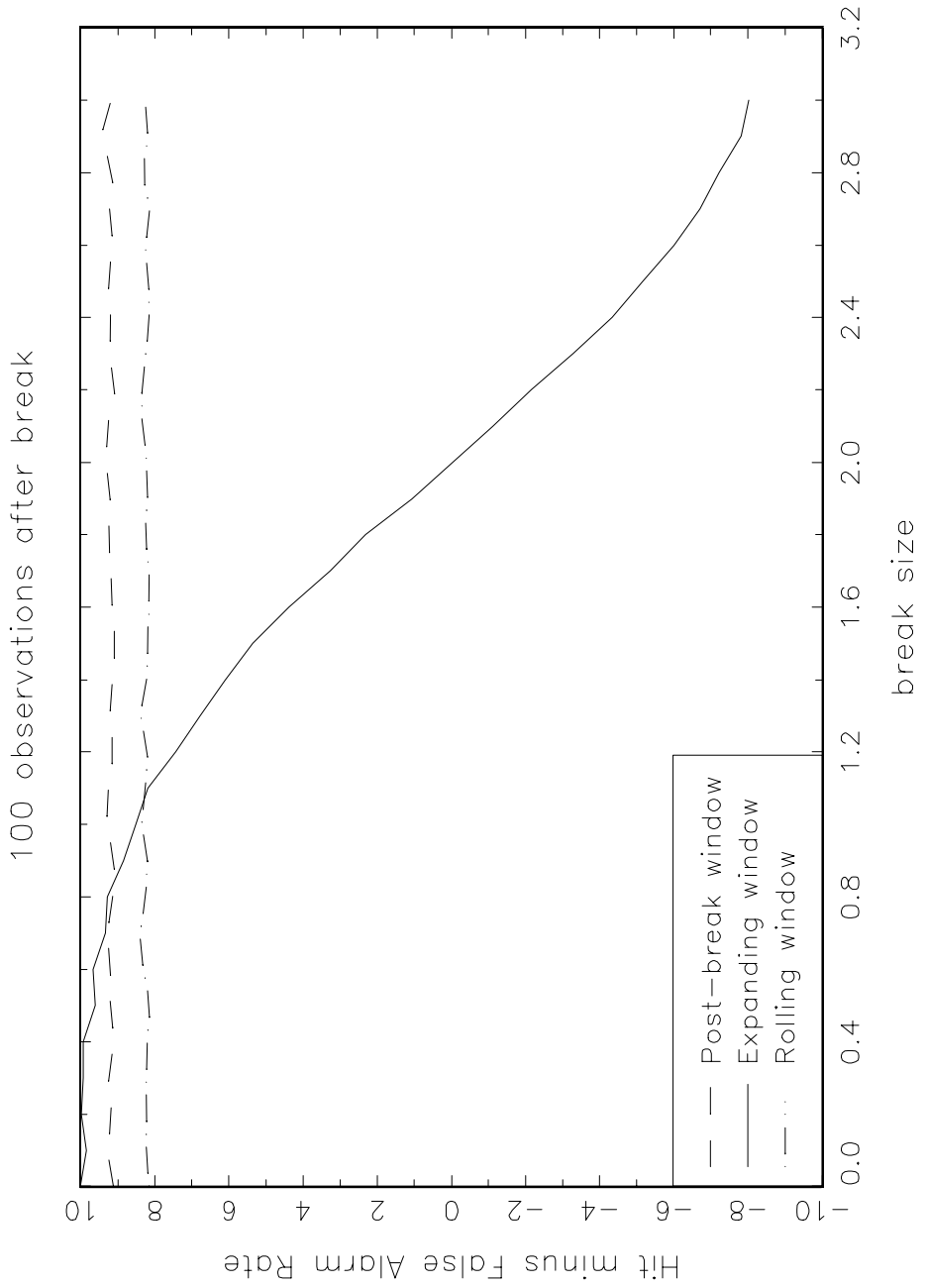

Figure 12: Hit minus false alarm rate (higher post-break variance)

Figures 10-12 assume that $\sigma_1 = 6, \sigma_2 = 9$ so the R^2 value of the prediction equation is very low. Consequently, $H - F$ is only about 4% for the post-break method based on 10 observations, although this rises to 6% and 9% when 25 and 100 post-break data points are available, respectively. For break sizes up to around 0.8, it is better to use the expanding window. The intuition for this finding is that it is difficult to precisely estimate the parameters of the forecasting model in the presence of very high levels of noise as under this scenario. The bias in the forecasts due to using pre-break data is therefore more than made up for by reduced parameter estimation uncertainty provided that the break size (and thus the bias) is not too large.

We conclude from these results that it is important to account for parameter instability in directional forecasting models. If the size of the discrete changes in the underlying parameters is small, it may not be detectable and forecasters may be best served by simply using an expanding window. As the break size grows larger, however, precision in directional forecasting requires a real time breakpoint monitoring procedure that attempts to identify the most recent break without too much delay and drops data prior to the break. The comparative performance of such a procedure relative to that of the popular rolling window approach will of course depend on the length of the rolling window, the size and frequency of breaks and on how easily and quickly breaks can be detected.

5. **Conclusion**

Many economic and financial time series undergo sudden, large breaks reflecting institutional changes, regime switches or breakdowns in market mechanisms. It is perhaps not surprising that many forecasting models appear to be unstable and there may be good theoretical reasons for these empirical findings. Timmermann and Granger (2002) argue that certain forms of predictability in financial returns are likely to self-destruct as a consequence of the efficient market hypothesis once the predictable patterns are publicized and sufficient capital is allocated towards exploiting them.

Forecasting series subject to structural breaks 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.

We considered both the case with a pure break in the mean as well as cases with a simultaneous break in the mean and variance of the forecasting model. While the latter case is more complicated, it is intuitively clear how a break in the variance affects our results. If the post-break variance is lower than the pre-break variance, the new data after the break is less noisy than the pre-break data and the predictive R^2 higher. This will lead to better performance of the post-break estimation window compared to rolling and expanding window methods. Conversely, if the post-break variance is higher than the pre-break variance, it will generally be more worthwhile to accept some bias in the parameter estimates of the forecasting model in exchange for lower parameter estimation uncertainty. This will benefit expanding and rolling window approaches that typically use pre-break data over a pure post-break estimation approach.

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. When a reversed ordered cusum squared method is used to determine the time of the most recent break, the percentage of correctly predicted signs of a monthly return forecasting model rises from 61.6% (based on an expanding window) to 65.7%. The results presented in this paper help to interpret such empirical findings.

Appendix

Proof of proposition 1

The moments of θ_m are key to the distribution of $\widehat{y}_{T+1}$ under different window sizes. To derive these, first note that

$$\theta_m = \frac{\sum_{t=m}^{T_1} x_{t-1}^2}{\sum_{t=m}^{T_1} x_{t-1}^2 + \sum_{t=T_1+1}^{T} x_{t-1}^2},$$

and under $x_t/\omega \sim IIN\ (\mu_x/\omega, 1)$ we have

$$\theta_m \overset{d}{\sim} \frac{\chi^2_{\nu_1}(\lambda_1)}{\chi^2_{\nu_1}(\lambda_1) + \chi^2_{\nu_2}(\lambda_2)},$$

where $\chi^2_{\nu_i}(\lambda_i)$ is distributed as a non-central chi–squared with ν_i degrees of freedom and the non-centrality parameter $\lambda_i = v_i(\mu_x/\omega)^2$. Recall that $\nu_1 = T_1 - m + 1$ and $\nu_2 = T - T_1$. Hence θ_m has a doubly non-central beta distribution with parameters $\nu_1/2$ and $\nu_2/2$ and the non-centrality parameters λ_1 and λ_2.[5] Approximating each of the non-central χ^2 variables in θ_m and using Patnaik's approximation (Patnaik (1949)), we have

$$\theta_m \overset{appr}{\sim} \left(\frac{\frac{\nu_1}{2} + 2\lambda_1}{\frac{\nu_1}{2} + \lambda_1}\right)\left(\frac{\frac{\nu_2}{2} + \lambda_2}{\frac{\nu_2}{2} + 2\lambda_2}\right) Beta(f_1, f_2),$$

where

$$f_i = \frac{\left(\frac{\nu_i}{2} + \lambda_i\right)^2}{\frac{\nu_i}{2} + 2\lambda_i}.$$

But noting that $\lambda_i = \nu_i(\mu_x/\omega)^2$ it is readily seen that $f_i = \nu_i k$ where

$$k = \frac{\left[1 + 2(\mu_x/\omega)^2\right]^2}{2 + 8(\mu_x/\omega)^2}, \tag{13}$$

and

$$\left(\frac{\frac{\nu_1}{2} + 2\lambda_1}{\frac{\nu_1}{2} + \lambda_1}\right)\left(\frac{\frac{\nu_2}{2} + \lambda_2}{\frac{\nu_2}{2} + 2\lambda_2}\right) = 1.$$

[5] See, for example, Johnson and Kotz (1970) pages 197-198.

Hence $\theta_m \overset{appr}{\sim} Beta(f_1, f_2)$ and $E(\theta_m)$ and $E\left(\theta_m^2\right)$ can be directly obtained from the moments of the (central) beta distribution:

$$\begin{aligned} E(\theta_m) &\approx \frac{f_1}{f_1+f_2} = \frac{\nu_1}{\nu_1+\nu_2} = \frac{\nu_1}{\nu}, \\ E\left(\theta_m^2\right) &\approx \frac{f_1(f_1+1)}{(f_1+f_2)(f_1+f_2+1)}, \\ &= \frac{k\nu_1(1+k\nu_1)}{k(\nu_1+\nu_2)\left[k(\nu_1+\nu_2)+1\right]}, \\ &= \left(\frac{\nu_1}{\nu}\right)\frac{(1+k\nu_1)}{(1+k\nu)}. \end{aligned}$$

Under the assumptions stated in Proposition 1, the conditional distribution of $\widehat{y}_{T+1}$ given $X_T = (x_T, x_{T-1}, ...)$ is

$$E(\widehat{y}_{T+1}|\mathbf{X}_T) = \beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T, \tag{14}$$

or, unconditionally,

$$E(\widehat{y}_{T+1}) = \beta_2 \mu_x + (\beta_1 - \beta_2)E(\theta_m x_T).$$

Also since θ_m depends on $x_1, .., x_{T-1}$, and the x_t's are IIN, then θ_m and x_T are independently distributed

$$E(\theta_m x_T) = \left(\frac{\nu_1}{\nu}\right)\mu_x,$$

and

$$E(\widehat{y}_{T+1}) = \beta_2\mu_x + \left(\frac{\nu_1}{\nu}\right)(\beta_1 - \beta_2)\mu_x.$$

Now consider the second moments of $(y_{T+1}, \widehat{y}_{T+1})$. It is easily seen that

$$V(y_{T+1}) = \beta_2^2\omega^2 + \sigma^2.$$

To derive $V(\widehat{y}_{T+1})$ we note that

$$V(\widehat{y}_{T+1}) = E[V(\widehat{y}_{T+1}|\mathbf{X}_T)] + V\left[E(\widehat{y}_{T+1}|\mathbf{X}_T)\right]. \tag{15}$$

Under $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the conditional variance of $\widehat{y}_{T+1}$ is given by

$$V(\widehat{y}_{T+1}|\mathbf{X}_T) = \frac{\sigma^2 x_T^2}{\sum_{t=m}^{T} x_{t-1}^2}.$$

Therefore, using the expression for $E\left(\widehat{y}_{T+1}|\mathbf{X}_T\right)$ given by (14) we have

$$h^2 = V(\widehat{y}_{T+1}) = \sigma^2 E\left(\frac{x_T^2}{\sum_{t=m}^{T} x_{t-1}^2}\right) + V\left[\beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T\right]. \tag{16}$$

But

$$\begin{aligned} V\left[\beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T\right] &= \omega^2 + (\beta_1 - \beta_2)^2 E[(\theta_m x_T - \mu_x \frac{\nu_1}{\nu})^2] \\ &\quad +2\beta_2(\beta_1 - \beta_2)E[(x_T - \mu_x)(\theta_m x_T - \mu_x \frac{\nu_1}{\nu})], \end{aligned} \tag{17}$$

and after some algebra yields

$$\begin{aligned} V\left[\beta_2 x_T + (\beta_1 - \beta_2)\theta_m x_T\right] &= \beta_2^2\omega^2 + (\frac{\nu_1}{\nu})(\beta_1 - \beta_2)^2\omega^2\psi \\ &\quad +2(\frac{\nu_1}{\nu})\beta_2(\beta_1 - \beta_2)\omega^2. \end{aligned} \tag{18}$$

where

$$\psi = \frac{(1 + k\nu_1) + \left(\frac{\nu_2}{\nu}\right)(\mu_x/\omega)^2}{1 + k\nu}. \tag{19}$$

To evaluate $E\left(x_T^2/\sum_{t=m}^{T} x_{t-1}^2\right)$, first note that

$$\frac{\nu x_T^2}{\sum_{t=m}^{T} x_{t-1}^2} = \frac{(x_T^2/\omega)^2}{\sum_{t=m}^{T}(x_{t-1}/\omega)^2/\nu}$$

and since $x_t/\omega \sim IIN\ (\mu_x/\omega, 1)$, then

$$\frac{\nu x_T^2}{\sum_{t=m}^{T} x_{t-1}^2} \overset{d}{\sim} \left[t_\nu(\delta, \lambda)\right]^2,$$

where $t_\nu(\delta, \lambda)$ is distributed as a double non-central t-distribution with ν degrees of freedom and the non-centrality parameters $\delta = \mu_x/\omega$ and $\lambda = \nu\left(\mu_x/\omega\right)^2$. Using results in Johnson and Kotz (1970, p. 214, eq. 25) we now have

$$\begin{aligned} \phi &= E\left(\frac{x_T^2}{\sum_{t=m}^{T} x_{t-1}^2}\right) = \left(\frac{1}{2}\right)\exp(-\frac{1}{2}\lambda)(1 + \delta^2) \times \\ &\quad \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j}{j!} \frac{\Gamma(\frac{1}{2}(\nu - 2) + j)}{\Gamma(\frac{1}{2}\nu + j)}. \end{aligned} \tag{20}$$

Substituting (20) and (18) in (16) we have

$$\begin{aligned} h^2 &= V(\widehat{y}_{T+1}) = \sigma^2\phi + \beta_2^2\omega^2 + \omega^2(\frac{\nu_1}{\nu})(\beta_1 - \beta_2)^2\psi \\ &\quad +2\beta_2(\beta_1 - \beta_2)\omega^2(\frac{\nu_1}{\nu}). \end{aligned}$$

Similarly

$$\begin{aligned} g &\equiv Cov(y_{T+1}, \widehat{y}_{T+1}) \\ &= E\left\{[\beta_2(x_T - \mu_x) + u_{T+1}]\left[\beta_2(x_T - \mu_x) + (\beta_1 - \beta_2)(\theta_m x_T - \frac{\mu_x \nu_1}{\nu}) + \frac{x_T \sum_{t=m}^{T} x_{t-1} u_t}{\sum_{t=m}^{T} x_{t-1}^2}\right]\right\} \\ &= \beta_2^2 \omega^2 + \beta_2(\beta_1 - \beta_2)\omega^2(\nu_1/\nu). \end{aligned}$$

Collecting the above results we now have

$$\begin{pmatrix} y_{T+1} \\ \widehat{y}_{T+1} \end{pmatrix} \sim F\left\{\begin{pmatrix} \beta_2 \mu_x \\ \beta_2 \mu_x + \left(\frac{\nu_1}{\nu}\right)(\beta_1 - \beta_2)\mu_x \end{pmatrix}, \mathbf{\Sigma}\right\},$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} \beta_2^2 \omega^2 + \sigma^2 & g \\ g & h^2 \end{pmatrix}.$$

Break in Error Variances

In the case where $\sigma_1^2 \neq \sigma_2^2$, the above distributional results hold, except that the unconditional variance of $\widehat{y}_{T+1}$ can no longer be derived analytically. In this case we have

$$V(\widehat{y}_{T+1}|\mathbf{X}_T) = x_T^2 \left(\frac{\sigma_1^2 \sum_{t=m}^{T_1} x_{t-1}^2 + \sigma_2^2 \sum_{t=T_1+1}^{T} x_{t-1}^2}{\left(\sum_{t=m}^{T} x_{t-1}^2\right)^2}\right),$$

and since x_T is distributed independently of x_{T-1}, x_{T-2},...we have

$$E\left[V(\widehat{y}_{T+1}|\mathbf{X}_T)\right] = \left(\mu_x^2 + \omega^2\right)\sigma_1^2 E\left[\frac{\sum_{t=m}^{T_1} x_{t-1}^2}{\left(\sum_{t=m}^{T} x_{t-1}^2\right)^2}\right] + \left(\mu_x^2 + \omega^2\right)\sigma_2^2 E\left[\frac{\sum_{t=T_1+1}^{T} x_{t-1}^2}{\left(\sum_{t=m}^{T} x_{t-1}^2\right)^2}\right]. \tag{21}$$

The remaing terms, namely $V\left[E\left(\widehat{y}_{T+1}|\mathbf{X}_T\right)\right]$, $Cov(y_{T+1}, \widehat{y}_{T+1})$, and $V(y_{T+1})$ are derived as before with σ^2 in the expression for these terms replaced by σ_2^2.

The expectations in (21) can be computed by stochastic simulation, noting that by assumption $x_t \sim IIN(\mu_x, \omega^2)$. Let $x_{t-1}^{(r)}$ be the $(t-1)$th observation in the rth draw and let r go from 1 to R, the number of replications. We then have the following approximation

$$E\left[\frac{\sum_{t=m}^{T_1} x_{t-1}^2}{\left(\sum_{t=m}^{T} x_{t-1}^2\right)^2}\right] \approx \frac{1}{R}\sum_{r=1}^{R}\left[\frac{\sum_{t=m}^{T_1}\left(x_{t-1}^{(r)}\right)^2}{\left(\sum_{t=m}^{T}\left(x_{t-1}^{(r)}\right)^2\right)^2}\right].$$

Similarly

$$E\left[\frac{\sum_{t=T_1+1}^{T} x_{t-1}^2}{\left(\sum_{t=m}^{T} x_{t-1}^2\right)^2}\right] \approx \frac{1}{R}\sum_{r=1}^{R}\left[\frac{\sum_{t=T_1+1}^{T}\left(x_{t-1}^{(r)}\right)^2}{\left(\sum_{t=m}^{T}\left(x_{t-1}^{(r)}\right)^2\right)^2}\right].$$

Proof of proposition 2

To compute the probabilities that $y_{T+1} > 0$, $\widehat{y}_{T+1} > 0$ as well as their joint probabilities, note that

$$\begin{aligned}
\mu_{mT} &= (\omega z + \mu_x)\left(\beta_2 + (\beta_1 - \beta_2)\frac{\chi_{v_1}^2(\lambda_1)}{\chi_{v_1}^2(\lambda_1) + \chi_{v_2}^2(\lambda_2)}\right), \\
\sigma_{mT}^2 &= (\omega z + \mu_x)^2 \frac{\sigma^2}{\chi_{v_1}^2(\lambda_1) + \chi_{v_2}^2(\lambda_2)},\ z \sim N(0,1).
\end{aligned}$$

Consider now Pr $(y_{T+1} > 0)$. We know from direct evaluation that

$$f(y_{T+1}) = \varphi\left(\frac{y_{T+1} - \beta_2\mu_x}{\sqrt{\beta_2^2\omega^2 + \sigma^2}}\right).$$

The same result can also be obtained using the mixture normal approach, namely

$$f(y_{T+1}) = \int_{x_T=-\infty}^{\infty} \varphi\left(\frac{y_{T+1} - \beta_2 x_T}{\sigma}\right)\varphi\left(\frac{x_T - \mu_x}{\omega}\right) dx_T,$$

and noting that

$$\int_{x_T=-\infty}^{\infty} \varphi\left(\frac{y_{T+1} - \beta_2 x_T}{\sigma}\right)\varphi\left(\frac{x_T - \mu_x}{\omega}\right) dx_T = \varphi\left(\frac{y_{T+1} - \beta_2\mu_x}{\sqrt{\beta_2^2\omega^2 + \sigma^2}}\right),$$

as required.

The remaining probabilities in the PT statistic can be similarly computed using the conditional joint probability distributions of $(y_{T+1}, \widehat{y}_{T+1})$. Using (2) and (3) it is easily seen that conditional on $\mathbf{X}_{mT}$, y_{T+1} and $\widehat{y}_{T+1}$ are independently and normally distributed, and

$$f(y_{T+1}, \widehat{y}_{T+1} \mid \mathbf{X}_{mT}) = \varphi\left(\frac{y_{T+1} - \beta_2 x_T}{\sigma}\right)\varphi\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right).$$

Hence, unconditionally

$$f(y_{T+1}, \widehat{y}_{T+1}) = \int_{z=-\infty}^{\infty} \int_{\chi^2_{v_1}=0}^{\infty} \int_{\chi^2_{v_2}=0}^{\infty} \varphi\left(\frac{y_{T+1} - \beta_2(\omega\ z + \mu_x)}{\sigma}\right) \times \varphi\left(\frac{\widehat{y}_{T+1} - \mu_{mT}}{\sigma_{mT}}\right) \varphi(z)\, f(\chi^2_{v_1}(\lambda_1)) f(\chi^2_{v_2}(\lambda_2)) dz\ d\chi^2_{v_1} d\chi^2_{v_2}.$$

and

$$\Pr(y_{T+1} > 0, \widehat{y}_{T+1} > 0) = \int_{y=0}^{\infty} \int_{\hat{y}=0}^{\infty} f(y, \widehat{y}) dy d\hat{y}.$$

Bibliography

Alogoskoufis, G.S. & Smith, R (1991). The Phillips Curve, the persistence of inflation, and the Lucas Critique: Evidence from exchange rate regimes, *American Economic Review,* 81, 1254-1275.

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point, *Econometrica,* 61, 821-856.

Andrews, D.W.K. & Ploberger, W. (1996). Optimal changepoint tests for normal linear regression.. *Journal of Econometrics,* 70, 9-38.

Ang A., & Bekaert, G. (2001). International asset allocation with regime shifts, *Review of Financial Studies,* forthcoming.

Ang, A. & Bekaert, G. (2002). Regime switches in interest rates, *Journal of Business and Economic Statistics*, 20, 163-182.

Bai, J. & Perron, P. (1998). Estimating and testing linear models with multiple structural changes, *Econometrica,* 66, 47-78.

Bai, J. & Perron, P. (2002). Computation and analysis of multiple structural change models, *Journal of Applied Econometrics*, forthcoming.

Brown, R.L., Durbin, J., & Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society, Series B*, 37, 149-192.

Chib, S (1998). Estimation and comparison of multiple change-point models, *Journal of Econometrics,* 86, 221-242.

Chow, G. (1960). Tests of equality between sets of coefficients in two linear regressions, *Econometrica,* 28, 591-605.

Chu, C-S J., Stinchcombe, M., & White, H. (1996). Monitoring structural change,. *Econometrica,* 64, 1045-1065.

Clements, M.P. & Hendry, D.F. (1998). *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

David, A. & Veronesi, P. (2001). Option prices with uncertain fundamentals: Theory and evidence on the dynamics of implied volatilities and over-/underreaction in the options market, mimeo, Federal Reserve Board of Governors and University of Chicago.

Driffill, J. & Sola, M. (1994). Testing the term Structure of Interest Rates from a Stationary Switching Regime VAR", *Journal of Economic Dynamics and Control*, 18, 601-628.

Elliott, G. & Mueller, U. (2002). Optimally testing general breaking processes in linear time series models, Manuscript, University of California, San Diego.

Garcia, R. &Perron, P. (1996). An analysis of the real interest rate under regime shifts, *Review of Economics and Statistics,* 78, 111-125.

Granger, C.W.J. & Pesaran, M.H. (2000). Economic and statistical measures of forecast accuracy, *Journal of Forecasting,* 19, 537-560.

Gray, S. (1996). Modeling the conditional distribution of interest rates as regime-switching process, *Journal of Financial Economics*, 42, 27-62.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-384.

Hansen, B.E. (1992). Tests for parameter instability in regressions with I(1) processes, *Journal of Business and Economic Statistics,* 10, 321-335.

Henriksson, R.D. & Merton, R.C. (1981). On market timing and investment ierformance. II. statistical procedures for evaluating forecasting skills, *Journal of Business*, 54, 513-533.

Inclan, C. & Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance,. *Journal of the American Statistical Association,* 89, 913-923.

Leitch, G. & Tanner, J.E. (1991). Economic forecast evaluation: profits versus the conventional error measures, *American Economic Review*, 81, 580-90.

Johnson, N.L. & Kotz, S. (1970). *Continuous Univariate Distributions - 2*, John Wiley.

Patnaik, P.B. (1949). The non-central χ^2- and $F-$distributions and their

applications, *Biometrika* 36, 202-232.

Ploberger, W., Kramer, W., & Kontrus, K. (1989). A new test for structural stability in the linear regression model, *Journal of Econometrics,* 40, 307-318.

Perez-Quiros, G. & Timmermann, A. (2000). Firm size and cyclical variations in stock returns, J*ournal of Finance*, 55, 1229-1262.

Pesaran, M.H. & Timmermann, A. (1992). A simple non-parametric test of predictive performance, *Journal of Business and Economic Statistics*, 10, 461-465.

Pesaran, M.H. & Timmermann, A. (2001). Market timing and return prediction under model instability, *Journal of Empirical Finance*, forthcoming.

Skouras, S. (1999). On discrete investment rules for financial markets", Ph.D. Thesis, European University Institute, Florence, Italy.

Stock, J.H. & Watson, M.W. (1996). Evidence on structural instability in macroeconomic time series relations,. *Journal of Business and Economic Statistics,* 14, 11-30.

Timmermann, A. (2001). Structural breaks, incomplete information, and stock prices, *Journal of Business and Economic Statistics,* 19, 299-314.

Timmermann, A. & Granger, C.W.J. (2002). Efficient market hypothesis and forecasting, *International Journal of Forecasting*, forthcoming.

CESifo Working Paper Series

(for full list see www.cesifo.de)

---

809 Florian Englmaier and Achim Wambach, Contracts and Inequity Aversion, November 2002

810 Sarbajit Sengupta, Delegating Recruitment under Asymmetric Information, December 2002

811 Rajshri Jayaraman, On the Partial Public Provision of a Private Good, December 2002

812 Stéphanie Stolz, Banking Supervision in Integrated Financial Markets: Implications for the EU, December 2002

813 Christian Keuschnigg, Taxation of a Venture Capitalist with a Portfolio of Firms, December 2002

814 Inés Macho-Stadler and David Pérez-Castrillo, Settlement in Tax Evasion Prosecution, December 2002

815 Rainer Niemann and Dirk Simons, Costs, Benefits, and Tax-induced Distortions of Stock Option Plans, December 2002

816 Jan-Egbert Sturm and Barry Williams, Deregulation, Entry of Foreign Banks and Bank Efficiency in Australia, December 2002

817 V. Anton Muscatelli, Patrizio Tirelli, and Carmine Trecroci, Monetary and Fiscal Policy Interactions over the Cycle: Some Empirical Evidence, December 2002

818 Claude Hillinger, A General Theory of Price and Quantity Aggregation and Welfare Measurement, December 2002

819 Erkki Koskela and Ronnie Schöb, Optimal Capital Taxation in Economies with Unionised and Competitive Labour Markets, December 2002

820 Sheilagh Ogilvie, Guilds, Efficiency, and Social Capital: Evidence from German Proto-Industry, December 2002

821 Hans Gersbach and Verena Liessem, Financing Democracy, December 2002

822 Costas Hadjiyiannis, Panos Hatzipanayotou, and Michael S. Michael, Optimal Tax Policies with Private-Public Clean-Up, Cross-Border Pollution and Capital Mobility, December 2002

823 François Ortalo-Magné and Sven Rady, Homeownership: Low Household Mobility, Volatile Housing Prices, High Income Dispersion, December 2002

824 Syed M. Ahsan and Panagiotis Tsigaris, Measuring the Social Discount Rate under Uncertainty: A Methodology and Application, December 2002

825 Kai A. Konrad, Altruism and Envy in Contests: An Evolutionarily Stable Symbiosis, December 2002

826 Robert S. Chirinko and Huntley Schaller, A Revealed Preference Approach to Understanding Corporate Governance Problems: Evidence from Canada, December 2002

827 Geir B. Asheim, Green National Accounting for Welfare and Sustainability: A Taxonomy of Assumptions and Results, December 2002

828 Andrea Gebauer, Chang Woon Nam, and Rüdiger Parsche, Lessons of the 1999 Abolition of Intra-EU Duty Free Sales for Eastern European EU Candidates, December 2002

829 Giacomo Corneo, Work and Television, December 2002

830 Vivek H. Dehejia and Yiagadeesen Samy, Trade and Labour Standards – Theory, New Empirical Evidence, and Policy Implications, December 2002

831 Geir B. Asheim and Wolfgang Buchholz, A General Approach to Welfare Measurement through National Income Accounting, December 2002

832 Aaron Tornell and Frank Westermann, The Credit Channel in Middle Income Countries, January 2003

833 Gebhard Flaig, Time Series Properties of the German Monthly Production Index, January 2003

834 Campbell Leith and Jim Malley, Estimated Open Economy New Keynesian Phillips Curves for the G7, January 2003

835 Burkhard Heer and Bernd Süssmuth, Inflation and Wealth Distribution, January 2003

836 Erkki Koskela and Leopold von Thadden, Optimal Factor Taxation under Wage Bargaining – A Dynamic Perspective, January 2003

837 Carola Grün and Stephan Klasen, Growth, Income Distribution, and Well-Being: Comparisons across Space and Time, January 2003

838 Robert S. Chirinko and Ulf von Kalckreuth, On the German Monetary Transmission Mechanism: Interest Rate and Credit Channels for Investment Spending, January 2003

839 Sascha O. Becker, Andrea Ichino, and Giovanni Peri, How Large is the "Brain Drain" from Italy?", January 2003

840 Albert Berry and John Serieux, All About the Giants: Probing the Influences on Growth and Income Inequality at the End of the 20th Century, January 2003

841 Robert Fenge and Martin Werding, Ageing and the Tax Implied in Public Pension Schemes: Simulations for Selected OECD Countries, January 2003

842 Robert Fenge and Martin Werding, Ageing and Fiscal Imbalances Across Generations: Concepts of Measurement, January 2003

843 Giovanni Andrea Cornia, The Impact of Liberalisation and Globalisation on Income Inequality in Developing and Transitional Economies, January 2003

844 Peter Fredriksson and Per Johansson, Program Evaluation and Random Program Starts, January 2003

845 Bernd Hayo and Matthias Wrede, Fiscal Equalisation: Principles and an Application to the European Union, January 2003

846 Syed M. Ahsan and Jaideep Oberoi, Inequality, Well-being and Institutions in Latin America and the Caribbean, January 2003

847 Chang Woon Nam and Doina Maria Radulescu, The Role of Tax Depreciation for Investment Decisions: A Comparison of European Transition Countries, January 2003

848 V. Bhaskar and Steinar Holden, Wage Differentiation via Subsidised General Training, January 2003

849 Paloma Lopez-Garcia, Labour Market Performance and Start-up Costs: OECD Evidence, January 2003

850 Christian Keuschnigg and Soren Bo Nielsen, Public Policy for Start-up Entrepreneurship with Venture Capital and Bank Finance, January 2003

851 Yin-Wong Cheung, Menzie D. Chinn, and Eiji Fujii, China, Hong Kong, and Taiwan: A Quantitative Assessment of Real and Financial Integration, January 2003

852 Gregory D. Hess, The Economic Welfare Cost of Conflict: An Empirical Assessment, February 2003

853 Douglas J. Cumming and Jeffrey G. MacIntosh, Comparative Venture Capital Governance. Private versus Labour Sponsored Venture Capital Funds, February 2003

854 Eckhard Janeba and John Douglas Wilson, Decentralization and International Tax Competition, February 2003

855 Tapio Palokangas, Capital Accumulation and Employment Cycles in a Model of Creative Destruction, February 2003

856 Brendan Walsh, When Unemployment Disappears: Ireland in the 1990s, February 2003

857 Luis H. R. Alvarez and Erkki Koskela, A General Approach to the Stochastic Rotation Problem with Amenity Valuation, February 2003

858 Christian Schultz, Strategic Campaigns and Redistributive Politics, February 2003

859 Ernst Fehr and Joseph Henrich, Is Strong Reciprocity a Maladaptation? On the Evolutionary Foundations of Human Altruism, February 2003

860 Haizhou Huang, Dalia Marin, and Chenggang Xu, Financial Crisis, Economic Recovery and Banking Development in Former Soviet Union Economies, February 2003

861 Pedro Cardoso and Bernard M.S. van Praag, How Sustainable Are Old-age Pensions in a Shrinking Population with Endogenous Labour Supply?, February 2003

862 Volker Meier, Efficient Transfer of Aging Provisions in Private Health Insurance, February 2003

863 Edward Castronova, Theory of the Avatar, February 2003

864 Robert S. Chirinko, Hans van Ees, Harry Garretsen, and Elmer Sterken, Investor Protections and Concentrated Ownership: Assessing Corporate Control Mechanisms in the Netherlands, February 2003

865 Bernard M.S. van Praag and Pedro Cardoso, The Mix Between Pay-as-you-go and Funded Pensions and what Demography has to do with it, February 2003

866 Ernst Fehr, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert G. Wagner, A Nation-Wide Laboratory. Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Survey, February 2003

867 Frank Heinemann, The Inflationary Impact of Wage Indexation, February 2003

868 Eytan Sheshinski, Bounded Rationality and Socially Optimal Limits on Choice in a Self-Selection Model, February 2003

869 M. Hashem Pesaran, Estimation and Inference in Large Heterogenous Panels with Cross Section Dependence, February 2003

870 Luis H. R. Alvarez and Erkki Koskela, On the Tree-Cutting Problem under Interest Rate and Forest Value Uncertainty, February 2003

871 Norbert Berthold and Rainer Fehn, Unemployment in Germany: Reasons and Remedies, February 2003

872 Clemens Fuest, Bernd Huber, and Philipp Tilleßen, Tax Policy and Entrepreneurship in the Presence of Asymmetric Information in Capital Markets, February 2003

873 Eytan Sheshinski, Optimum and Risk-Class Pricing of Annuities, February 2003

874 Willi Leibfritz, Paul O'Brien and Jean-Christophe Dumont, Effects of Immigration on Labour Markets and Government Budgets – An Overview, February 2003

875 M. Hashem Pesaran and Allan Timmermann, How Costly is it to Ignore Breaks when Forecasting the Direction of a Time Series?, February 2003