Real-Time Forecasting with Vector Autoregressions: Spurious Drift, Structural Change and Intercept-Correction

Ronald Bewley

The University of New South Wales, Sydney 2052, Australia

e-mail: r.bewley@unsw.edu.au

This draft: January 2001

JEL Classifications: C32, C53, E17

KEYWORDS: time series, forecast comparisons, VAR, BVAR

---

* The project was funded by an *Australian Research Council* grant (A10007132). This paper has benefited greatly from many useful discussions with, and comments from, David Hendry, whom I thank. The procedures developed in this paper were inspired by ‘Hans’ Theil’s repeated least squares and mixed estimation innovations. Sadly, Hans, my dear friend, mentor, and sometime coauthor passed away while I was writing this paper. I would like to acknowledge Hans’ support and encouragement over the years. I have also benefited from presentations at the UNSW time series group weekly workshops and the excellent research assistance provided by Glenn Anderson.

Abstract

A recent literature, dominated by David Hendry and Michael Clements, has emerged which emphasises the importance of the estimated constant term in Vector Autoregressive (VAR) forecasting models using nonstationary time series. A part of this research agenda has resulted in the resurrection of intercept-corrections, once popular with the large econometric models of the 1970s, in the VAR context owing to the common occurrence of intermittent structural breaks and the importance of the constant term in determining medium and longer-term forecasting trajectories. This procedure is further developed in this paper and a new Bayesian estimator is proposed. It is shown that, in many commonly occurring situations, the proposed estimator produces a substantial gain in mean squared forecast error over the alternatives of not intercept-correcting, or always correcting in the classical fashion.

1. Introduction

In its simplest form, a Vector Autoregression (VAR) is an unrestricted reduced form model that expresses each variable as a linear function of a constant and the lags of that and each other variable in the system. Since each equation in a VAR has the same regressors, they can be estimated separately by OLS. However, even in moderately sized systems with, say, six variables and four lags of each, and a constant term, there are 25 parameters to be estimated in each equation so that over-parameterisation has often been cited as the main cause of poor forecasting performance.

Since their introduction in the late 1970s, one strand of research has focussed on reducing the parameter space by imposing stochastic restrictions which imply that each variable follows a random walk with drift [Doan, Litterman and Sims (1984), Litterman (1980, 1986)]. These restrictions are justified by appealing to the common observation that most economic time series are integrated to order one, I(1). The combination of these priors with data-determined parameter estimates produces the variant commonly known as the BVAR, or Bayesian VAR, with the so-called Litterman priors after its originator. However, the estimation procedure is more properly known as mixed estimation [Theil and Goldberger (1961)] and it can trivially be implemented using OLS with a data augmentation dummy variable procedure [Robertson and Tallman (1999)].

The original BVAR models have become commonplace in the literature with their relative success usually being attributed to the general reduction in the parameter space [e.g. Artis and Zhang (1990), Holden and Broomhead (1990), Funke (1990), and Shoesmith (1995)]. In a dissenting note, Bewley (2000a) argues that this improvement is more likely due to the correction of the downward bias in the estimated unit roots (or near unit roots) that has been shown to increase with the number of variables in the system [Abadir, Hadri and Tzavalis (1999)]. More recently, Sims and Zha (1998), have further developed the BVAR methodology with priors that have more general applicability and an estimation procedure that is more in the true Bayesian spirit.

A separate line of enquiry has focussed on pretesting the time series for unit roots and, where appropriate, testing for cointegration in the spirit of Engle and Granger (1987) and Johansen (1988). While these methods typically result in only a minimal reduction in the parameter space, the imposition of unit roots have been shown to produce marked improvements in forecast accuracy over both an unrestricted VAR and a BVAR with Litterman priors [Bewley (2000a, 2000b)]. The main justification for this result is that forecasts of I(1) variables from Vector Error Correction (VEC) models rapidly approach linear time trends. Therefore, it is the precision of the estimate of the constant term, rather than short-run dynamics, that is the key issue in forecasting with I(1) time series [Clements and Hendry (1995, 1998b) and Bewley (2000b)].

Clements and Hendry's (1995, 1998b) taxonomy of the sources of forecast error shows that, if it is assumed that the data generating process (dgp) is stable over time, and a consistent estimator is used, there are only three important sources of forecast error in a VAR in levels with I(1) data: the failure to impose unit roots; overparameterisation arising from an unrestricted reduced form specification; and, importantly, the precision in estimation of the drift parameters. In the standard VAR framework, the constant terms are nonlinear functions of all of the drift parameters and all of the (many) coefficients on the lagged variables, so that it is a nontrivial matter to address the question of controlling the estimates of drift in that framework. However, Bewley (2000b) proposed a new representation which allows a simple, possibly Bayesian, direct estimation of the drift parameters in a VEC. Thus, in cases where some variables contain drift while others do not, the so-called mixed drift case, substantial improvements in forecast accuracy can be made by imposing exact, or stochastic, restrictions on the drift parameters.

Since the estimation of drift is of such importance in VEC models, Bewley and Yang (2000) developed a number of statistics to test for structural breaks specifically in that parameter vector. While system tests previously existed [Andrews (1993) and Andrews and Ploberger (1994)], these new tests allow for a simple variable-by-variable test for parameter constancy that are more appropriate when only some of the variables are subject to structural breaks in drift.

The methods introduced in this paper are designed to take a properly-specified model, that has been successfully subjected to appropriate diagnostic tests, and use that model in a real-time setting. In practice, structural breaks are commonplace and potentially an extremely important contributor to poor forecasting performance. As a result, Hendry (1996, 1997), Hendry and Clements (1994), and Clements and Hendry (1995, 1996, 1998a, 1998b) have suggested correcting the intercept at each forecast origin to realign the forecasts after a break has occurred within the sample. This correction is achieved by adding the most recent residual, or the average over the last few periods, to the one-step forecasts, or to the forecasts at all lead times.

While Clements and Hendry advocate intercept-corrections with VEC models, they argue that they have less merit in a VAR model in the differences (DVAR). Indeed, they consider a (non-intercept-corrected) DVAR, which results from a mis-specifying a VEC model, as a viable alternative to intercept-correcting a VEC model. In that sense, Clements and Hendry consider breaks in the long-run means of the equilibrium- (error-) correction terms as more important, and more likely, sources of forecast error than breaks in drift parameters. This view is supported by Eitrheim, Husbeø and Nymoen’s (1999) experimentation with a macroeconometric model of the Norwegian Central Bank.

While there is clear merit in the Clements and Hendry prescription, the results in Bewley (2000b) suggest that a fruitful approach might be to allow the drift parameters, rather than the constant terms in a VEC, to change in the recent past in both a classical and a mixed-estimation sense. While these methods can be applied to either VEC or DVAR models, the emphasis in this paper is on the latter. It is found that substantial improvements in forecast accuracy can be made by intercept-correcting DVARs.

The proposed intercept-correction methods are introduced in Section 2 for the general VEC model. The design of the simulation experiment is presented in Section 3 and the results are reported in Section 4. Conclusions are drawn in Section 4.

2. Intercept-Correction

Consider an n x 1 vector of I(1) times series, y_t, that can be represented by a VAR with p lags, VAR(p):

$$y_t = a + \sum_{i=1}^{p} A_i y_{t-i} + u_t \tag{1}$$

If $\Sigma A_i - I$ has rank $r > 0$, equation (1) can be written as VEC(p):

$$\Delta y_t = a + \alpha [\beta' y_{t-1}] + \sum_{i=1}^{p-1} B_i \Delta y_{t-i} + u_t \tag{2}$$

where $B_j = -\Sigma^p_{i=j+1} A_i$ and $\Sigma^p_{i=1} A_i - I = \alpha\beta'$, with α and β being n x r. If the time series are not cointegrated (r=0), then a VAR in Δy_t with p-1 lags is appropriate, DVAR(p-1).

Clements and Hendry (1998b) write equation (2) as

$$[\Delta y_t - \delta] = \alpha[\beta' y_{t-1} - \gamma] + \sum_{i=1}^{p-1} B_i [\Delta y_{t-i} - \delta] + u_t \tag{3}$$

where γ is the mean of the equilibrium-correction terms, $\beta' y_{t-1}$, and δ is the mean of Δy_t, the drift parameters. The forecasts of (3) approach linear time trends with slope δ. It follows from a comparison of equations (2) and (3) that the constant term, a, in equation (2) is a nonlinear function of the drift, δ, and the all of the B_j:

$$a = [(I - \sum_{i=1}^{p-1} B_i)\,\delta - \alpha\,\gamma] \tag{4}$$

Bewley (2000b) proposed an alternative representation of equation (2) or (3) that enables the direct estimation of the drift parameter vector, δ. This is achieved by applying the Bewley (1979) transformation to equation (3):

$$\Delta y_t = \delta + \zeta\,[\beta' y_{t-1} - \gamma] + \sum_{i=0}^{p-2} C_i\,\Delta^2 y_{t-i} + v_t \tag{5}$$

where the C_i are nonlinear functions of all of the B_j and ζ is a nonlinear function of all of the B_j and α. Since each equation in (5) is exactly identified, two stage least squares (2SLS), conditional upon a super-consistent estimate of β, and using a constant term, the equilibrium-correction terms and the p-1 lags of Δy_t as instruments, is equivalent to indirect least squares implying an exact relationship between the estimates of the three basic forms of a VEC: equations (2), (3) and (5). Bewley (2000b) introduced equation (5) to enable some of the drift parameters to be set to zero with exact or stochastic restrictions in the mixed-drift situation. Equation (5) is referred to as the mixed-drift VAR (MDVAR).

When restrictions are placed on the MDVAR, the equations are no longer exactly identified and a system estimation approach, such as three stage least squares (3SLS), or iterated 3SLS, is appropriate.

If it is assumed that the means of the equilibrium-correction terms are constant over time, but that all of the drift parameters have changed in a recent time period, an appropriately-defined dummy variable can be added to each equation of (5) or, equivalently, equations (2) or (3). This is one notion of intercept-correction. If only a subset of the drift parameters has changed, it is not possible to impose linear restrictions on the dummy variable augmentation to (2) or (3). However, this is not the case in the MDVAR, and so possible gains might be had by exactly or stochastically restricting the dummy variable parameters in that framework.

On the other hand, if the long-run means of the equilibrium correction terms, ζ, vary, it is not possible to distinguish between changes in the two long run means, δ and ζ with a single dummy variable for each equation. Importantly, changes in any one element of ζ implies that all of the dummy variable coefficients are non-zero in either the VEC or MDVAR. In that sense, it is not clear that there is an advantage to pursuing the latter when equilibrium-correction terms are included and are possibly subject to structural breaks. The notion that the equilibrium-correction terms do not break, but the individual series do, has been termed co-breaking by Hendry and discussed in Clements and Hendry (1998b) and elsewhere.

Since it may be the case that only a subset of the time series have been subjected to a structural break, the mixed-break case, but that this is not known to the model builder, consideration of a Bayesian intercept-correction method, which introduces a dummy to equation in (5) but shrinks the coefficients to zero, may prove fruitful. Such a procedure has two potential advantages. First, shrinking the intercept-correction towards zero implies that the stochastically-restricted MDVAR is less likely to over-react to noise. Second, those series without breaks would have dummy variable coefficients with large standard errors resulting in restricted estimates that would be attracted towards zero more so than for series that did break. When a similarly defined Bayesian intercept-correction is applied to a standard DVAR or VEC, the mixed-break case implies all of the dummy variables coefficients are typically non-zero so that this differential shrinking to zero cannot easily be exploited.

In order to implement this stochastic restriction on a DVAR or VEC, it is useful to consider a the i^{th} equation of a DVAR as a standard regression equation

$$w = X\pi + \varepsilon \tag{6}$$

using obvious notation (conditional upon the equilibrium-correction terms having been defined) and assuming the first column of X is the dummy variable introduced for intercept-correction. The stochastic restriction can be written as

$$r = R\pi + \nu \tag{7}$$

where r = 0 and R is a row vector of zeroes except for the first element which is unity.

It is necessary to prespecify the variance of ν which controls the so-called tightness of the prior; $Var(\nu) = \lambda^2$. Assuming $Var(\varepsilon) = \sigma^2 I$, stacking (6) and (7) produces the augmented model

$$\begin{bmatrix} w \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \pi + \begin{bmatrix} \varepsilon \\ \nu \end{bmatrix} \tag{8}$$

which might be estimated by GLS since σ is not necessarily equally to λ. Alternatively, equation (7) can be multiplied through by $\hat{\sigma}/\lambda$ and this equation stacked with equation (6)

$$\begin{bmatrix} w \\ r \end{bmatrix} = \begin{bmatrix} X \\ R^* \end{bmatrix} \pi + \begin{bmatrix} \varepsilon \\ \nu^* \end{bmatrix} \tag{8}$$

where the first element of R^* is $\hat{\sigma}/\lambda$ and the remaining elements are zero. The variance of ε and ν^* are now approximately the same. Thus, equation (8) can be estimated by OLS; the so-called dummy variable augmentation procedure. In keeping with the BVAR literature, this estimate of σ is taken from a p-1^{th} order autoregression of $\Delta y_{i,t}$.

When exact or stochastic restrictions are applied to the MDVAR, certain modifications are necessary. The correlation of X with ε requires X to be replaced by its least squares predictions $\hat{X}$ in typical 2SLS fashion. However, as $\lambda \to 0$, the resultant estimates do not approach those that would have been computed if the dummy variables had been excluded from the system, as would be the case if this were applied in the DVAR/VEC situation. This difference arises because the instruments used in creating $\hat{X}$ include the dummy variable which is then being 'excluded' by the restrictions. The difference might be expected to be very small when the restriction is true but, when a significant break has occurred, sizeable differences can emerge in parameter estimates and forecast accuracy. To rectify this problem, the first-stage regressions can similarly be augmented with the dummy variable procedure to shrink the first-stage dummy variable coefficients to zero. That is, the $T+1^{th}$ observations are all zero except for that corresponding to the dummy variable which is set to $\hat{\sigma}/\lambda$. In this way, the restricted estimates approach those from not including an intercept-correction as $\lambda \to 0$ and the unrestricted estimates, including an intercept-correction term, as $\lambda \to \infty$. While $\hat{\sigma}$ is taken from autoregressions of $\Delta y_{i,t}$ in the case of a DVAR/VEC, $\hat{\sigma}$ is taken from the Bewley transformation of that autoregression in the MDVAR.

3. Monte Carlo Design

In a series of papers, Hendry, and Clements and Hendry, have argued that intercept-corrections to VEC models perform on a par with DVARs without corrections and that corrections to the latter tend to reduce forecast performance due to an over-reaction to noise. Since the analysis of Section 2 suggests that there might be greater potential to make improvements in MDVARs than DVARs or VEC models, the ensuing experiment has been designed to investigate the merits, if any, of intercept-correcting an MDVAR when there is no cointegration. Thus, consider a two-equation mixed-break dgp with T observations and no cointegration:

$$\Delta y_t = \delta_1 + \phi_1(\Delta y_{t-1} - \delta_1) + \phi_2(\Delta x_{t-1} - \delta_2 - \delta^* D_{t-1}) + \varepsilon_{1,t} \quad (9)$$

$$\Delta x_t = \delta_2 + \delta^* D_t + \phi_3(\Delta x_{t-1} - \delta_2 - \delta^* D_{t-1}) + \varepsilon_{2,t} \quad (10)$$

where ε_i is n.i.d.(0,1) and $E(\varepsilon_1\varepsilon_2) = 0$. D_t is the dummy variable that introduces the structural break in the drift of x_t but not y_t. Note that this dummy is lagged when it operates on Δx_{t-1} to align it with the timing of the break.

Equivalently, (9) and (10) can be expressed as

$$\Delta y_t = \gamma_1 + \mu_1 D_{t-1} + \phi_1 \Delta y_{t-1} + \phi_2 \Delta x_{t-1} + \varepsilon_{1,t} \quad (9')$$

$$\Delta x_t = \gamma_2 + \delta^* D_t + \mu_2 D_{t-1} + \phi_3 \Delta x_{t-1} + \varepsilon_{2,t} \quad (10')$$

where $\gamma_1 = (1-\phi_1)\delta_1 - \phi_2\delta_2$; $\gamma_2 = (1-\phi_3)\delta_2$; $\mu_1 = -\phi_2\delta^*$; $\mu_2 = -\phi_3\delta^*$. Thus, both constants shift in a DVAR as shown in (9′) and (10′), providing $\phi_2 \neq 0$, but there is no observable shift in y_t because the shift in x_t is offset by the change in the drift term in equation (9). However, in a real-time forecasting situation, there is a delay of one period before the constant term in (9′) changes as, in the first period of the change, there is a single unit element in D_t corresponding to the last observation and, hence, in that special case $D_{t-1} = 0$.

It is assumed that $D_t = 0$ if $t < T-q$ and $D_t = 1$ if $t \geq q$. That is, the break in drift is permanent (and continues throughout the forecasting period). Moreover, q is also assumed to be sufficiently small as to make the shift difficult to detect. Any previous changes in drift are assumed to have been modelled. Following the construction of the Andrews and Ploberger (1994) and Bewley and Yang (2000) Wald tests, breaks are assumed not to have occurred at the ends of the sample, say within 15% of the end of the sample. This serves as a guide to the value of q. Two values of q are considered, $q = 0$ and $q = 4$. Clearly, even if a significant shock were to be detected in the last observation, corresponding to $q = 0$, it would be impossible to test whether the shock was temporary or permanent. In two cases corresponding to $q = 0$ and 4, it is assumed in estimation that the location of the hypothesised break is known. In a third case, the assumed location, q^*, is set to 4 while the true value is $q = 0$. This serves two purposes. In reality, the location of a break would not be known and secondly, it is consistent with Clements and Hendry's approach of averaging recent errors so that over-reaction to shocks is less prevalent.

Two basic dgp's are considered. The first, referred to as the 'diagonal dgp', has $\phi_1 = \phi_3 = 0.5$ and $\phi_2 = 0$. This implies that the constant term in the first equation of the DVAR does not depend on the size of the break in the second, δ^*; that is, there is no 'transmitted' structural break. In the second dgp, $\phi_1 = \phi_3 = 0.5$ so that the speed of adjustment to equilibrium is unchanged but $\phi_2 = -0.8$, implying that both constant terms in the DVAR depend upon δ^*; that is, there is a 'transmitted' structural break in the first equation but not the first variable. Three values of δ^* are considered: 1, 2 and 4 and these compare to $Var(\Delta y_t) = Var(\Delta x_t) = 1/(1\text{-}0.5^2) = 4/3$. Thus, $\delta^* = 4$ is unlikely to be encountered in many applications but is included to assess the impact of the proposed methods in extrema.

Since none of the forecasting comparisons depend on the drift parameters, both were arbitrarily set equal to one. These values, together with the other parameters chosen for the dgp, produce time series that visually have the characteristics of many macroeconomic time series. That is, they exhibit strong drift, but deviations from a linear trend are pronounced. A stylised example for a single replication of the triangular dgp is given in Figure 1 for $T = 85$ and $\delta^* = 0,1,2,4$ with a break occurring at observation 71. In the case of the diagonal dgp, the y series has the same dgp as x but without a break.

Each experiment is based on 5,000 replications and T = 75, 150 with the forecast comparisons being based on the 1- and 10-step ahead mean squared forecast error (msfe). In each case, the true lag length of one in first-differences was chosen so as to abstract from any forecast inaccuracy due to selecting the wrong lag length. The DVAR was estimated with no correction and with an unrestricted intercept-correction. The forecasts from these are necessarily the same as those from the comparable MDVARs. In addition, each form was estimated in its Bayesian form for 41 different values of λ [= 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.5, 3.0, 3.5, 4.0 4.5, 5.0, 6.0, 7.0. 8.0, 9.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 50.0, 75.0, 100.0]

4. Results

The 1-step and 10-step-ahead msfe's for the DVAR with no intercept-correction ($\lambda = 0$) and with an unrestricted intercept-correction ($\lambda \to \infty$) together with the Bayesian variant of the DVAR and MDVAR that produced the smallest msfe for that estimator and that equation, are presented in Tables 1 and 2 for the case T = 75.

When there is no transmitted drift, as there is in the diagonal dgp, the first variable has no structural break and, therefore, it would be expected that the no intercept-correction model would produce the smallest msfe. On the other hand, the second variable does exhibit a break, but it does not follow that an unrestricted intercept-correction is to be preferred. This is the essence of the Clements and Hendry view that DVARs should not be intercept-corrected. However, as the size of the break increases, unrestricted correction would become increasingly preferable.

From Table 1 it can be noted that both the 1-step and 10-step forecasts with the diagonal dgp do indeed have lower msfe's when $\lambda = 0$ than when $\lambda \rightarrow \infty$, and that the estimated λ is either zero or very small in the case of the DVAR. However, very small gains in msfe are apparent in the MDVAR with most of the estimated values of λ being in the approximate range of 0.2 - 0.5 for 1-step ahead forecasts. Small values of λ are also estimated at the longer lead time.

Turning to Table 2, it can be noted that the ranking of the forecast performance for the second variable, and the diagonal dgp for the cases of $\lambda = 0$ and $\lambda \rightarrow \infty$, depends upon the size of the break, δ^*, the number of periods for which the break has been in operation, q, and the number of periods for which the dummy variable assumes a change, q^*. Not surprisingly, it is better to intercept-correct (in the unrestricted sense) when δ^* is large, except when the break occurred in the last observation and residual-averaging ($q^* > 0$) does not occur. Whether or not λ is estimated in the DVAR or the MDVAR form, the msfe is never lower in either extreme and, sometimes, the gains from the Bayesian procedure are very substantial and greater than that achieved by residual-averaging.

In the case of the diagonal dgp and variable 2, it can be also noted from Table 2 that the estimated values of λ are typically larger than those in Table 1 reflecting the existence of the structural change and its accommodation with an intercept-correction. Most values for the DVAR are in the range 0.5 - 1.1 while that range is approximately 1.5 - 5.0 for the MDVAR.

The results for the triangular dgp are quite different from those with a diagonal dgp. In the triangular case, both equations in the DVAR experience a structural break while only the second breaks in the MDVAR form. For variable 1, unrestricted intercept-correction is preferable to no correction in the case $q^* = q = 4$ but the reverse is true for the longer lead time. In most cases the Bayesian DVAR performed similarly to the no-correction DVAR and substantial gains can be found in the MDVAR variant. The estimated values of λ are mostly in the range of 0.7 - 3.0 for the Bayesian MDVAR.

Not surprisingly, some of the biggest gains in msfe can be found in the triangular dgp and variable 2 and the Bayesian MDVAR is far superior to the Bayesian DVAR in many of the cases. Indeed, the Bayesian MDVAR produces the smallest (or equivalent) msfe in all but one of the 72 cases reported in Tables 1 and 2.

Two problems arise from the comparisons apparent in Tables 1 and 2. First, the value of λ is not restricted to be the same in the two equations. Unless the model builder has different information about the likelihood of a break in each variable, a common value of λ should be used. Second, it is not possible to ascertain the sensitivity of the msfe for λ in the neighbourhood of the minimum.

The former problem can be addressed by combining the msfe for the two equations for a common value of λ. Naturally, the results depend upon the weights attached to each series. In Table 3, the two msfe's are simply added reflecting the similar trends, scale and variability of the two time series. In no case is the Bayesian DVAR, the unrestricted intercept-correction, or the no-correction DVAR, preferred to the Bayesian MDVAR and, in some cases, the gains are very large indeed. The biggest gains are made when the break is very large, $\delta^* = 4$, and the dgp is triangular. However, even when $\delta^* = 1$, $q = q^* = 4$, and the dgp is triangular, the gain in msfe for 10-step ahead forecasts of the Bayesian MDVAR is 20% over not correcting and 50% over intercept correction.

The sensitivity of the results to changes in λ is considered in Figures 2 - 9 for a break of $\delta^* = 2$ and $q = q_* = 4$. In each case there are three pairs of lines corresponding to variable 1, with the suffix (1), variable 2, with the suffix (2), and the sum of two msfe's, with the suffix (1+2). Within each pair, one line refers to the Bayesian DVAR and the other to the Bayesian MDVAR. A log scale is used for λ to allow for additional detail at smaller values of λ.

It can be seen by comparing all eight figures that there is a pronounced minimum sum of msfe's for the Bayesian MDVAR. In some cases, such as in Figure 7, the scale of the graph masks the full extent of the gain for that model. These minima also occur for variable two but it is the combination of the flatness of the curve for variable 1 over a wide range of λ, and the minimum of the curve for variable 2, that make the overall gain so pronounced.

The Bayesian MDVAR produces a lower sum of msfe, particularly at the longer lead time, and for a reasonably wide range of λ. In each case, an arbitrarily selected value of 1 would have produced a forecasting model preferable to either an unrestricted correction or no correction. While a more extensive set of experiments is warranted, the results of this design are sufficiently encouraging to explore the use of the Bayesian intercept-correction MDVAR in practice.

The results for T = 150 are not presented, but are available from http:// economics.web.unsw.edu.au/people/rbewley, owing to their similarity with those given here. In essence, the additional observations improve the precision of the estimates of the basic DVAR but there are still only 1 (q = 0) or 5 (q = 4) observations in the new regime after the break. In that sense, the forecast relativities of the model do not depend upon sample size.

5. Conclusions

A new strategy has been proposed for estimating and testing VAR forecasting models, including the possibility of structural change, in a series of papers: Bewley

(2000a, 2000b) and Bewley and Yang (2000), and the present paper. The essence of the approach is to focus attention on the drift parameters. Spurious drift, arising from poorly determined drift estimates of variables that would not be expected to trend, cause problems of forecast credibility and, while relevant restrictions could be placed upon a standard VEC or DVAR model, the nonlinearity of the restrictions make such estimation cumbersome and prone to optimisation problems. The proposed framework for estimating VARs - the mixed drift VAR (MDVAR) - facilitates this estimation and has been shown to enable more accurate, and more credible forecasts to be produced (Bewley 2000b) using either exact or stochastic restrictions.

The computational time to estimate the proposed parameterisation is trivial owing to there being an analytical solution. However, there is much scope for experimenting with different priors when stochastic restrictions are being contemplated but even this is quite feasible. For example 5,000 replications, 41 sets of priors, 75 to 150 observations, and two equations took only 13 minutes with a Pentium II processor and a program written in FORTRAN.

In a natural extension to Bewley (2000b), Bewley and Yang proposed a set of statistics to test for structural change specifically in the drift parameters. The MDVAR framework simplifies the allocation of any structural change in the system to attribute it to possibly a subset of the variables. The present paper considers possible structural change in the most recent observations and an automatic Bayesian procedure for adapting to structural change in drift without overacting to noise is proposed. It has been found that, contrary to previous research (Clements and Hendry), significant gains in mean squared forecast error can be made by intercept-correcting a VAR in differences. Indeed, when there is a significant degree of causality in the system, and structural change is confined to a subset of the variables, substantial gains in forecast accuracy can occur with the new estimator.

While the usual caveat of drawing conclusions from Monte Carlo experiments is at least as applicable here as elsewhere, the results are sufficiently encouraging to warrant further investigation. In particular, work is progressing on extending these results to discriminating between structural breaks in drift and in the long-run means of equilibrium correction terms.

6. References

Abadir, K.M., Hadri, K., Tzavalis, E., 1999. The influence of VAR dimensions on estimator biases. Econometrica 67, 63-181.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 825-856.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

Artis, M.J., Zhang, W., 1990. BVAR forecasts for the G-7. International Journal of Forecasting 6, 349-62.

Bewley, R.A., 1979. The direct estimation of the equilibrium response in a linear dynamic model. Economics Letters 3, 357-361.

Bewley, R., 2000a. Forecast accuracy, coefficient bias and Bayesian vector autoregressions. Mathematics and Computers in Simulation, forthcoming.

Bewley, R., 2000b. Controlling spurious drift in macroeconomic forecasting models. Unpublished mimeo (The University of New South Wales, Sydney, Australia).

Bewley, R., Yang, M., 2000. Testing for structural breaks in the long-run means of VARs.

Unpublished mimeo (The University of New South Wales, Sydney, Australia).

Clements, M.P., Hendry, D.F., 1995. Forecasting in cointegrated systems. Journal of Applied Econometrics 10, 127-146.

Clements, M.P., Hendry, D.F., 1996. Intercept corrections and structural change. Journal of Applied Econometrics 11, 475-494.

Clements, M.P., Hendry, D.F., 1998a. Forecasting economic processes. International Journal of Forecasting 14, 111-131.

Clements, M.P., Hendry, D.F. 1998b. Forecasting Economic Time Series. Cambridge University Press: Cambridge, UK.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projections using realistic prior distributions. Econometric Reviews 3, 1-100.

Eitrheim, Ø., Husbeø, T.A., Nymoen, R., 1999. Equilibrium correction vs. differencing in macroeconomic forecasting. Economic Modelling 16, 515-544.

Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction: representations, estimation and testing. Econometrica 55, 252-276.

Funke, M., 1990. Assessing forecast accuracy of monthly vector autoregressive models. International Journal of Forecasting 6, 363-378.

Hendry, D.F., 1996. On the constancy of time series econometric equations. Economic and Social Review 27, 401-422.

Hendry, D.F., 1997. The econometrics of macroeconomic forecasting. Economic Journal 107, 1330-1357.

Hendry, D.F., Clements, M.P., 1994. On a theory of intercept corrections in macroeconomic forecasting. In Holly, S. (Ed.), Money, Inflation and Employment: Essays in Honour of James Ball. Edward Elgar: Aldershot, pp. 160-182.

Holden, K., Broomhead, A., 1990. An examination of vector autoregressive forecasts for the U.K. economy. International Journal of Forecasting 6, 11-23.

Johansen, S., 1988. Statistical analysis of cointegrating vectors. Journal of Economic Dynamics and Control 12, 231-254.

Litterman, R.B., 1980. A Bayesian procedure for forecasting with vector autoregressions, unpublished mimeo (Massachusetts Institute of Technology: Cambridge, Mass.).

Litterman, R.B., 1986. Forecasting with Bayesian vector autoregressions - five years of experience. Journal of Business and Economic Statistics 4, 25-37.

Robertson, J.C., Tallman, E.W., 1999., Vector autoregressions: forecasting and reality. Federal Reserve Bank of Atlanta Economic Review, First Quarter, 4-18.

Shoesmith, G.L., 1995. Multiple cointegrating vectors, error correction, and forecasting with Litterman's model. International Journal of Forecasting 11, 557-567.

Sims, C.A., Zha, T.A., 1998. Bayesian methods for dynamic multivariate models. International Economic Review 39, 949-968.

Theil, H., Goldberger, A.S., 1961. On pure and mixed statistical estimation in economics. International Economic Review 2, 65-78.

Table 1: Mean Squared Forecast Errors for Variable 1 (no break in drift)

				DVAR	DVAR		MDVAR		DVAR
δ^*	q^*	q	dgp	λ=0	λ	mse	λ	mse	$\lambda=\infty$
1-step									
1	0	0	diag	1.05	0.1	1.05	0.7	1.04	2.01
2	0	0	diag	1.09	0.1	1.09	0.8	1.07	2.05
4	0	0	diag	1.23	0.1	1.23	1.0	1.20	2.19
1	0	4	diag	1.05	0.1	1.05	0.4	1.05	1.24
2	0	4	diag	1.09	0.1	1.09	0.4	1.08	1.28
4	0	4	diag	1.23	0.1	1.23	0.5	1.23	1.43
1	4	4	diag	1.04	0.1	1.04	0.5	1.04	1.22
2	4	4	diag	1.07	0.1	1.07	0.5	1.06	1.22
4	4	4	diag	1.13	0.1	1.13	0.5	1.13	1.23
1	0	0	tri	1.73	0.1	1.73	1.8	1.38	2.73
2	0	0	tri	3.78	0.1	3.78	3.0	1.68	4.78
4	0	0	tri	11.88	0.0	11.88	3.5	3.10	12.90
1	0	4	tri	1.73	0.0	1.73	1.0	1.67	1.93
2	0	4	tri	3.78	0.0	3.78	1.7	3.25	4.00
4	0	4	tri	11.88	0.0	11.88	2.0	9.68	12.16
1	4	4	tri	1.58	1.5	1.24	2.5	1.16	1.25
2	4	4	tri	2.77	∞	1.33	3.5	1.22	1.33
4	4	4	tri	4.50	∞	1.57	∞	1.57	1.57
10-step									
1	0	0	diag	39.19	0.0	39.19	0.2	39.17	375.87
2	0	0	diag	39.73	0.0	39.73	0.3	39.70	405.19
4	0	0	diag	42.08	0.0	42.08	0.3	41.99	524.02
1	0	4	diag	39.19	0.0	39.19	0.2	39.16	100.67
2	0	4	diag	39.73	0.0	39.73	0.2	39.69	103.31
4	0	4	diag	42.08	0.0	42.08	0.2	42.02	114.32
1	4	4	diag	39.18	0.0	39.18	0.2	39.15	100.15
2	4	4	diag	39.75	0.0	39.75	0.2	39.73	101.01
4	4	4	diag	43.01	0.0	43.01	0.2	43.00	103.83
1	0	0	tri	122.88	0.0	122.88	1.2	119.81	1516.56
2	0	0	tri	160.11	0.0	160.11	1.5	143.18	3338.25
4	0	0	tri	311.47	0.0	311.47	1.5	251.77	10616.73
1	0	4	tri	122.88	0.0	122.88	0.7	121.79	316.26
2	0	4	tri	160.11	0.0	160.11	0.8	155.70	480.62
4	0	4	tri	311.47	0.0	311.47	0.8	296.54	1148.62
1	4	4	tri	121.26	0.0	121.26	1.1	112.95	281.68
2	4	4	tri	153.00	0.0	153.00	1.3	121.95	336.38
4	4	4	tri	268.94	0.0	268.94	1.6	192.50	497.74

Table 2: Mean Squared Forecast Errors for Variable 2 (break in drift)

δ^*	q^*	q	dgp	DVAR λ=0	DVAR λ	DVAR mse	MDVAR λ	MDVAR mse	DVAR $\lambda=\infty$
1-step									
1	0	0	diag	1.35	0.6	1.22	1.6	1.19	2.24
2	0	0	diag	2.16	0.9	1.32	2.5	1.24	2.92
4	0	0	diag	5.47	1.1	1.47	3.0	1.21	5.62
1	0	4	diag	1.35	0.4	1.32	0.9	1.32	1.42
2	0	4	diag	2.16	∞	1.82	6.0	1.82	1.82
4	0	4	diag	5.47	∞	3.45	∞	3.45	3.45
1	4	4	diag	1.29	0.5	1.17	1.3	1.16	1.29
2	4	4	diag	1.78	0.8	1.21	2.5	1.17	1.32
4	4	4	diag	2.66	1.2	1.24	4.5	1.15	1.39
1	0	0	tri	1.34	0.6	1.21	1.5	1.19	2.26
2	0	0	tri	2.13	0.9	1.32	2.5	1.25	2.96
4	0	0	tri	5.36	1.1	1.49	3.0	1.21	5.76
1	0	4	tri	1.34	0.4	1.31	1.0	1.30	1.40
2	0	4	tri	2.13	∞	1.78	5.0	1.78	1.78
4	0	4	tri	5.36	∞	3.34	∞	3.34	3.34
1	4	4	tri	1.28	0.5	1.16	1.4	1.15	1.27
2	4	4	tri	1.75	0.8	1.20	2.5	1.16	1.30
4	4	4	tri	2.51	1.2	1.22	4.5	1.14	1.38
10-step									
1	0	0	diag	114.37	0.6	78.97	1.5	76.39	427.98
2	0	0	diag	344.50	0.9	110.47	2.5	95.47	659.90
4	0	0	diag	1265.87	1.1	152.36	3.0	86.38	1587.59
1	0	4	diag	114.37	0.3	106.20	0.9	106.27	135.42
2	0	4	diag	344.50	∞	235.24	7.0	234.90	235.24
4	0	4	diag	1265.87	∞	629.90	∞	629.90	629.90
1	4	4	diag	103.84	0.5	66.69	1.3	64.01	102.80
2	4	4	diag	288.39	0.8	78.56	2.5	67.71	108.21
4	4	4	diag	849.37	1.1	86.82	5.0	60.73	123.44
1	0	0	tri	115.13	0.6	78.56	1.6	76.52	412.05
2	0	0	tri	347.46	0.9	110.48	2.5	94.38	635.81
4	0	0	tri	1277.64	1.1	159.79	3.5	89.84	1529.30
1	0	4	tri	115.13	0.4	106.45	0.9	106.78	134.29
2	0	4	tri	347.46	∞	238.43	7.0	238.13	238.43
4	0	4	tri	1277.64	∞	651.79	∞	651.79	651.79
1	4	4	tri	105.50	0.5	66.03	1.4	63.56	100.58
2	4	4	tri	298.05	0.8	77.46	2.5	66.79	106.22
4	4	4	tri	922.26	1.2	85.30	5.0	60.00	123.31

Table 3: Sum of Mean Squared Forecast Errors for Variables 1 and 2

				DVAR	DVAR		MDVAR		DVAR
δ^*	q^*	q	dgp	λ=0	λ	mse	λ	mse	$\lambda=\infty$
1-step									
1	0	0	diag	2.40	0.5	2.31	1.3	2.26	4.25
2	0	0	diag	3.25	0.8	2.55	2.0	2.44	4.97
4	0	0	diag	6.70	1.1	2.97	3.0	2.67	7.82
1	0	4	diag	2.40	0.2	2.38	0.6	2.37	2.66
2	0	4	diag	3.25	0.7	3.04	1.8	3.03	3.10
4	0	4	diag	6.70	∞	4.87	∞	4.87	4.87
1	4	4	diag	2.33	0.4	2.25	1.0	2.23	2.51
2	4	4	diag	2.85	0.7	2.35	2.0	2.31	2.54
4	4	4	diag	3.79	1.1	2.43	4.5	2.37	2.62
1	0	0	tri	3.07	0.5	2.97	1.7	2.57	4.99
2	0	0	tri	5.91	0.8	5.18	2.5	2.93	7.74
4	0	0	tri	17.23	1.1	13.53	3.5	4.36	18.66
1	0	4	tri	3.07	0.3	3.05	1.0	2.97	3.33
2	0	4	tri	5.91	0.7	5.68	1.9	5.07	5.78
4	0	4	tri	17.23	∞	15.50	2.5	13.46	15.50
1	4	4	tri	2.86	0.9	2.44	1.7	2.32	2.53
2	4	4	tri	4.51	3.5	2.63	3.0	2.39	2.63
4	4	4	tri	7.01	∞	2.95	7.0	2.86	2.95
10-step									
1	0	0	diag	153.57	0.4	132.40	1.1	129.50	803.85
2	0	0	diag	384.23	0.7	201.87	1.9	186.21	1065.09
4	0	0	diag	1307.95	1.0	316.36	3.0	246.07	2111.60
1	0	4	diag	153.57	0.2	149.47	0.6	149.33	236.09
2	0	4	diag	384.23	0.7	318.83	1.8	317.46	338.55
4	0	4	diag	1307.95	∞	744.23	∞	744.23	744.23
1	4	4	diag	143.02	0.4	118.49	1.0	116.24	202.95
2	4	4	diag	328.14	0.6	149.87	2.0	142.32	209.22
4	4	4	diag	892.38	1.0	174.36	4.5	156.44	227.27
1	0	0	tri	238.01	0.2	234.99	1.5	196.96	1928.61
2	0	0	tri	507.57	0.3	486.34	2.0	252.89	3974.06
4	0	0	tri	1589.10	0.4	1493.40	2.5	493.36	12146.03
1	0	4	tri	238.01	0.1	237.50	0.9	228.74	450.55
2	0	4	tri	507.57	0.2	499.65	1.5	427.80	719.05
4	0	4	tri	1589.10	0.4	1523.64	2.0	1203.70	1800.42
1	4	4	tri	226.76	0.3	215.36	1.3	177.24	382.26
2	4	4	tri	451.04	0.5	336.04	1.9	202.49	442.59
4	4	4	tri	1191.20	0.9	544.45	3.5	358.20	621.05

Figure 1: Stylised Time Series with Breaks

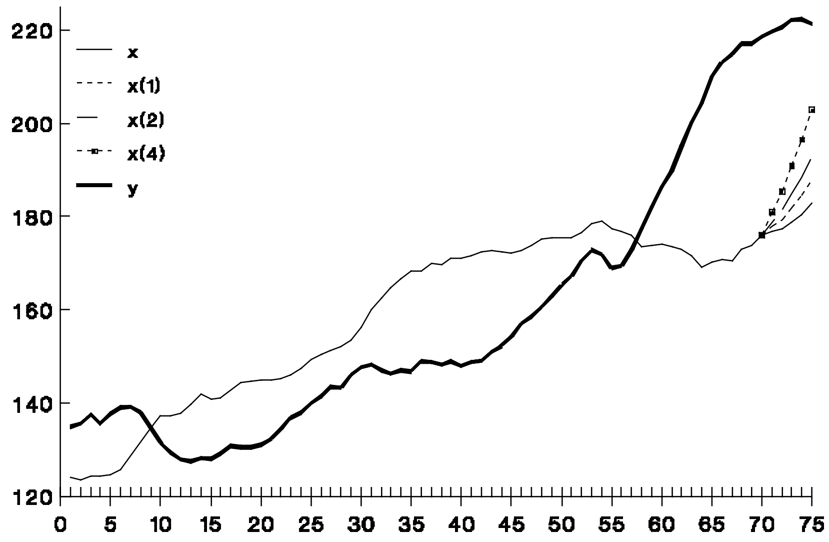

Figure 2: 1-step-ahead MSFE for Diagonal DGP and $\delta^* = 2$, $q^* = q = 0$

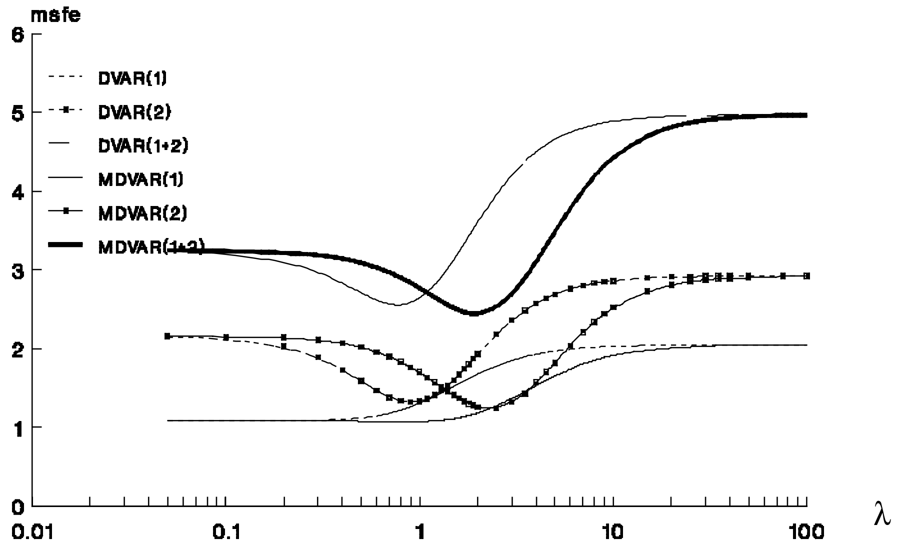

Figure 3: 10-step-ahead MSFE for Diagonal DGP and $\delta^{*} = 2$, $q^{*} = q = 0$

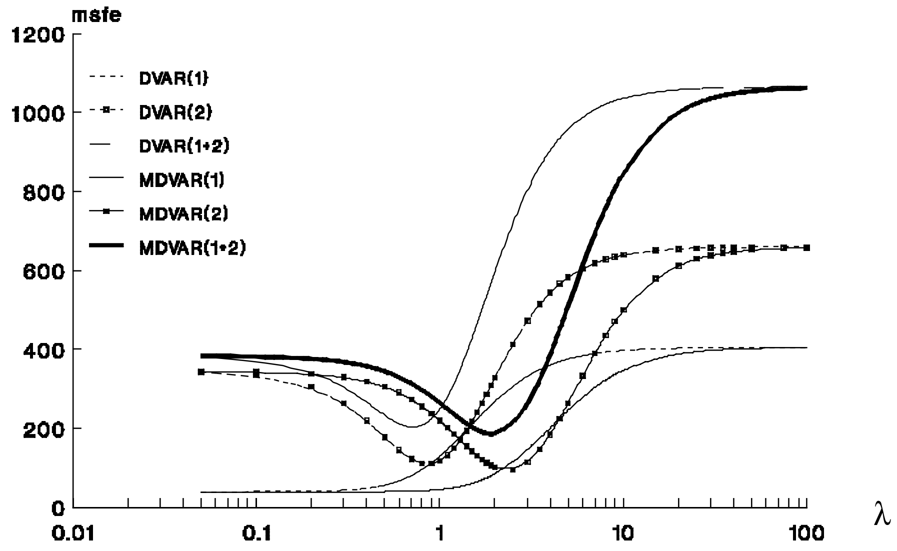

Figure 4: 1-step-ahead MSFE for Diagonal DGP and $\delta^* = 2$, $q^* = q = 4$

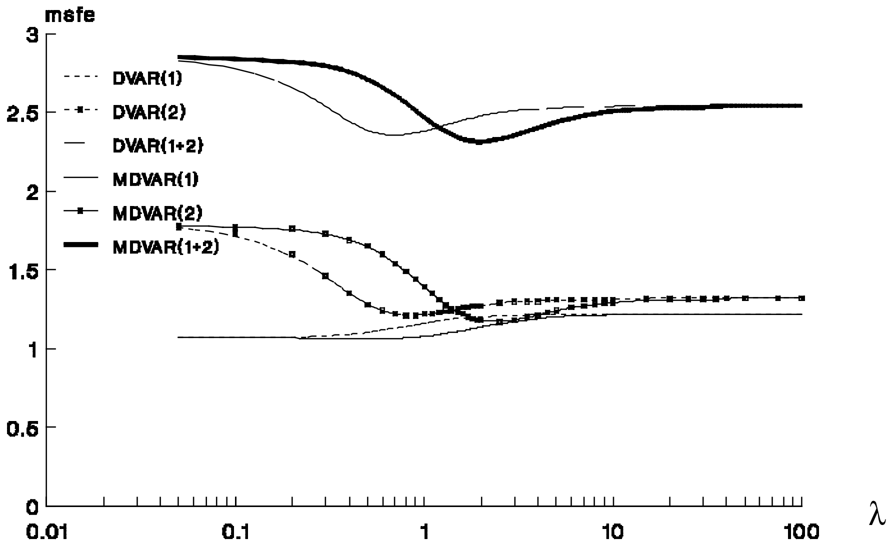

Figure 5: 10-step-ahead MSFE for Diagonal DGP and $\delta^* = 2$, $q^* = q = 4$

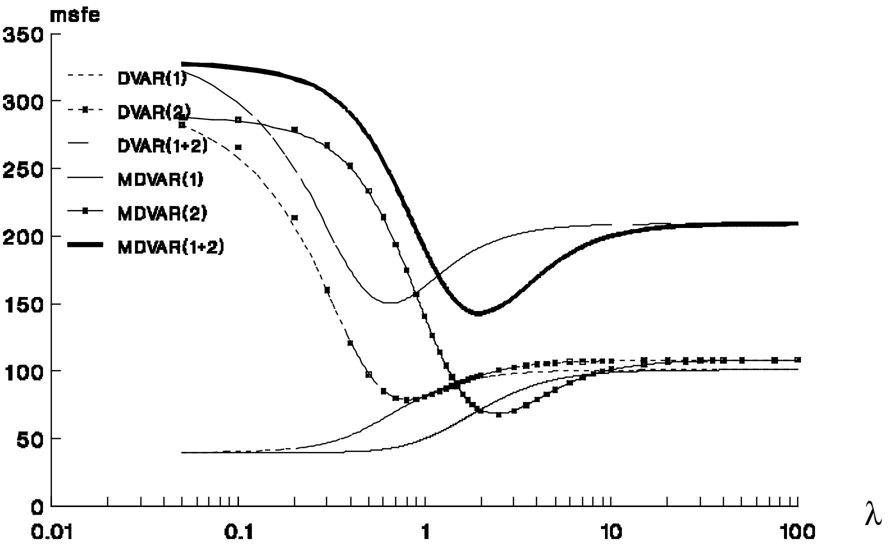

Figure 6: 1-step-ahead MSFE for Triangular DGP and $\delta^* = 2$, $q^* = q = 0$

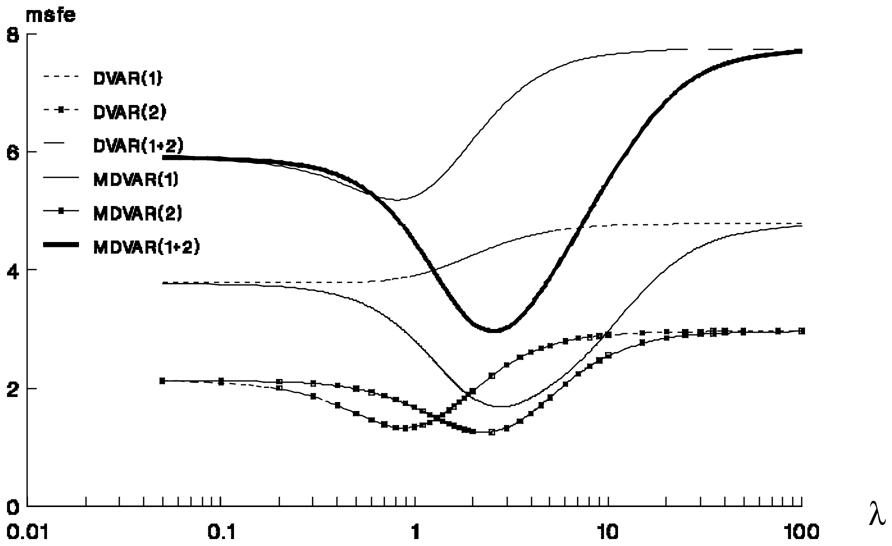

Figure 7: 10-step-ahead MSFE for Triangular DGP and $\delta^{*} = 2$, $q^{*} = q = 0$

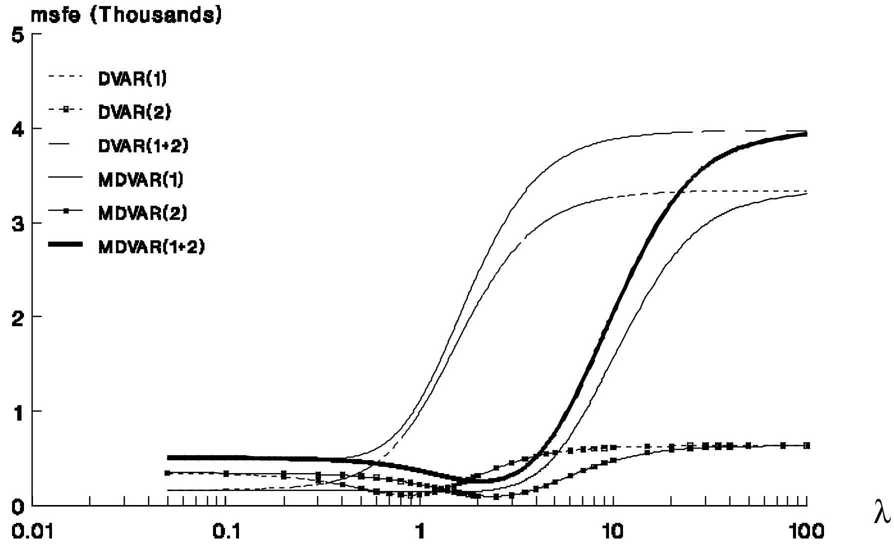

Figure 8: 1-step-ahead MSFE for Triangular DGP and $\delta^{*} = 2$, $q^{*} = q = 4$

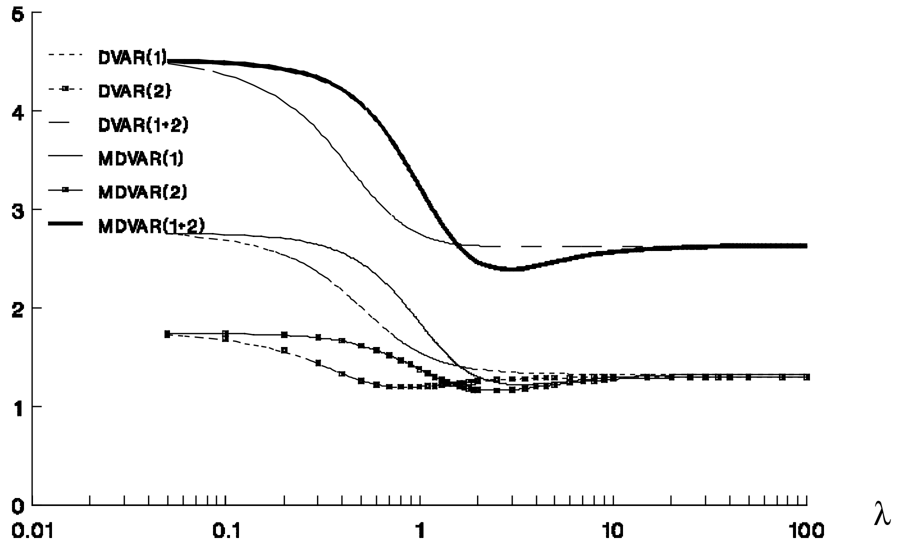

Figure 9: 10-step-ahead MSFE for Triangular DGP and $\delta^{*} = 2$, $q^{*} = q = 4$