Document de travail

A COMPARISON OF MULTI-STEP GDP FORECASTS FOR SOUTH AFRICA

Une comparaison de prévisions multi-étapes du PIB sud-africain

N° 2004-13
Décembre 2004

GUILLAUME CHEVILLON*
OFCE et Economics Department, Oxford University

*e-mail : guillaume.chevillon@sciences-po.fr
Adresse : OFCE, 69 quai d'Orsay, 75340 Paris cedex 07, France.

I am grateful to David Hendry, Janine Aron, John Muellbauer and Michael Clements for helpful comments and suggestions.

Observatoire Français des Conjonctures Économiques
69, Quai d'Orsay 75340 Paris Cedex 07
Tel : 01 44 18 54 00 Fax : 01 45 56 06 15
E-mail: ofce@ofce.sciences-po.fr Web: http://www.ofce.sciences-po.fr

Abstract

To forecast at several, say h, periods into the future, a modeller faces two techniques: iterating one-step ahead forecasts (the IMS technique) or directly modeling the relation between observations separated by an h-period interval and using it for forecasting (DMS forecasting). It is known that structural breaks, unit-root non-stationarity and residual autocorrelation benefit DMS accuracy in finite samples, all of which occuring when modelling the South African GDP over the last thirty years. This paper analyzes the forecasting properties of the model developped by Aron and Muellbauer (2002) and compares them with that of 30 derived or competing models. We find that the GDP of South Africa is best forecast, 4 quarters ahead, using the technique developed by these authors and its variants as derived in the present paper. Rankings of other models vary over time and it is difficult to recommend one of them as a rule in this exercise.

Keywords: Multi-step forecasting, Structural breaks, Forecast comparisons.

JEL *Classification*: C22, C53, E3.

Résumé

Pour prévoir à un horizon postérieur à la prochaine observation, par exemple à l'horizon h, un modélisateur a le choix entre deux techniques : soit itérer les prévisions à une étape (la technique IMS) ou modéliser directement le lien entre des observations séparées par un intervalle de h périodes et l'utiliser pour la prévision (technique DMS). Il est connu que les ruptures structurelles, la non-stationnarité causée par une racine unitaire ou l'autocorrélation des résidus favorisent la prévision par la méthode DMS dans des échantillons de taille finie. Tous ces facteurs sont présents lorsqu'on s'attaque la modélisation du PIB Sud-Africain au cours des trente dernières annes. Cet article analyse les propriétés prévisionnelles du modèle développé par Aron et Muellbauer (2002) et les compare avec celles de 30 modèles concurrents. Nous trouvons que le PIB Sud-Africain est mieux prévu (un horizon de quatre trimestres) en utilisant la méthode obtenue par ces auteurs et certaines de ses variantes développées ici. Les performances des autres modèles varient au cours du temps et il semble difficile d'en recommander certains de manière générale.

Mots-clefs: Prévision multi-étapes, Ruptures structurelles, Comparaison de prévisions.

Codes **JEL**: C22, C53, E3.

Contents

When a forecaster uses a model with a given periodicity but wishes to forecast at several, say $h > 1$, periods into the future, she is faced with a choice between iterating one-step ahead forecasts or *directly* modelling the relation between the end-of-sample observation and its hth successor in order to forecast the latter. It has been shown in empirical examples and theoretical analyses that, in finite samples, the second technique (direct multi-step or DMS) can prove more accurate than the former (iterated multi-step of IMS) when the data are non-stationary—be it from stochastic or deterministic origin—or if the model is misspecified for the error process, see inter alia Chevillon and Hendry (2004), Chevillon (2004a) and Chevillon (2004b). We purpose here to assess these analytical results by observing the performance of the model which was developed by Aron and Muellbauer (2002) for forecasting the South African GDP. In a recent comparison of 171 U.S. macroeconomic time series, Marcellino, Stock, and Watson (2004) have exhibited little empirical forecast accuracy gain from the use of DMS over IMS. These authors were, however, focusing on stationary series (differences of the series, in the case of integrated processes). By contrast Aron and Muellbauer (2002) established a relationship which allowed them, in spite of the many breaks experienced by the economy, to forecast the annual change in the quarterly GDP series over the last twenty years. Because it directly targets the endogenous variable several periods ahead of the forecast origin, their method belongs to the class of direct multi-step forecasts. Unfortunately, Aron and Muellbauer (AM henceforth) were unable to assess the accuracy of their model by comparing it to alternative techniques. This is what we attempt below.

The plan of this paper is as follows. First, we review the South African context and the model developed by Aron and Muellbauer (2002). In section 2, we derive alternative—IMS and DMS—models from multivariate analyses. We, then proceed to a general comparison of the forecasting techniques, including others which were shown by Clements and Hendry (1999) to perform generally well. As the South African economy has undergone a lot of breaks in the last thirty years, our strategy consists in estimating the models recursively so that we observe the evolution of ex-ante forecast accuracy over time. Finally, section 4 concludes.

1 Forecasting the South African GDP

1.1 Thirty years of breaks

South Africa has undergone a profound transition in the last thirty years and, hence, any model of its economy would be subjected to frequently occurring breaks. From 1976 and with the government's policy of Apartheid, the country began suffering from increasing international isolation which culminated, between late 1985 and the 'free' elections of 1994, by a period of almost no access to international capital. These factors, combined to the high degree of reliance of the economy to mineral exports might explain some of the shocks and large variations observed in the economic variables (see the articles by Aron and Muellbauer for an extended analysis of the South African context).

Following Aron and Muellbauer, we can distinguish three main monetary regimes since the 1960s. Until the late 1970s, there existed quantitative controls on interest rates and credit and the main criteria used for monetary policy were liquid asset ratios, while the corrective effect of interest rates was largely neglected by the regulatory authorities. Financial liberalization and transition towards a more flexible, cash-reserves based, system took place over the first half of the 1980s. From 1986 onwards, the monetary authorities made use of the discount rate to influence the interbank overnight refinancing market in order to achieved pre-declared monetary targets. The credit growth which followed the financial liberalization soon lessened the usefulness of monetary targets, thus leading, from 1998, to a new regime. The South African Reserve Bank (SARB) now offers some amount at the daily tender for repurchase transactions, thus signalling its preferences on short-term interest rates via an auction mechanism. Since early 2000, inflation targeting was reinstated as part of the medium-term monetary objectives.

Following the developments of monetary policy, the foreign exchange market experienced various regimes pointing towards greater flexibility. From US dollar or pound Sterling pegs combined with restrictions on resident and nonresident capital flows, the system moved, in 1979, to a regime of dual currency. Most non-resident transactions operated at the floating 'financial' exchange rate, while a 'commercial' rate was instated and announced in line with market forces. The latter became market determined in 1983 and the dual rates were soon re-unified. A debt crisis and the collapse of the Rand provoked a return to the dual currency system after 1985. In 1995, unification of the dual currency was initiated under a managed rate which has become fully floating at the introduction of inflation targeting.

We reproduce Table 1 of AM (table 1, here) where they present the various regimes experienced by the South African economy.

1.2 The AM model

Aron and Muellbauer developed a model for forecasting the annual change of the South African real GDP (Y, or y in log) via a solved-out equilibrium correction equation which depends on a set of variables $\{X_i\}$:

$$\Delta_4 y_{t+4} = \gamma \left(\alpha_0 + \mu_t + \sum_{i=1}^{n} \alpha_i X_{i,t} - y_t \right) + \sum_{i=1}^{n} \sum_{s=1}^{k} \beta_{i,s} \Delta X_{i,t-s} + \epsilon_t,$$

where ϵ_t is assumed white noise, but may be modelled by some moving average component and μ_t is a smooth stochastic trend which aims to capture the underlying production capacity of the

Table 1: Monetary Policy and Exchange Rate Policy Regimes (Aron and Muellbauer, 2002)

Period	**Monetary Policy Regimes**
1960–81	Liquid asset ratio-based system with quantitative controls on interest rates and credit
1981–85	Mixed system during transition
1986–98	Cost of cash reserves-based system with pre-announced M3 targets.
1998–99	Daily tenders of liquidity through repurchase transactions (repo system), plus pre-annouced M3 targets and tagets for core inflation
2000–	Repo system with inflation targetting
Period	**Exchange Rate Policy Regime**
1961(1)–71(2)	Pegged to fixed pound Sterling
1971(3)–74(2)	Pegged in episodes to floating US dollar/pound Sterling
1974(3)–75(2)	'Controlled independent float': devaluations every few weeks
1975(3)–79(1)	Fixed Regime: pegged to the US dollar
1979(2)–82(4)	Dual foreign exchange system: controlled floating commercial rand and floating financial rand
1983(1)–85(3)	Unification to a controlled floating rand
1985(4)–95(1)	Return to a dual system
1995(2)–	Unification to a controlled floating rand

economy. The stochastic trend is defined as in Harvey (1993) and Harvey and Jaeger (1993) by:

$$\mu_t = \mu_{t-1} + \delta_t + \eta_{1t},$$
$$\delta_t = \delta_{t-1} + \eta_{2t},$$

where η_{1t} and η_{2t} follow independent white noise processes. The stochastic properties of μ_t depend on those of (η_{1t}, η_{2t}) as follows:

$$\mathsf{Var}[\eta_{1t}] = 0 \quad \Rightarrow \mu_t \text{ is a smooth } \mathsf{I}(2) \text{ trend};$$
$$\mathsf{Var}[\eta_{2t}] = 0 \quad \Rightarrow \mu_t \text{ follows a random walk (with drift if } \delta_0 \neq 0).$$

It is the former case which we model here; it can be estimated with a Kalman filter as in the STAMP package (Koopman, Harvey, Doornik, and Shephard, 2000). AM study several sets of variables for $\{X_i\}$ and they settle with a few. They find that their equation is stable over the various regimes using as regressors the real prime interest rate (*RPRIME*) and its annual change, the ratio of current account surplus to current GDP (*RCASUR*), the government surplus to GDP ratio (*RGSUR*), the long-term growth of terms of trade (*TOT*), the difference in a financial liberalization indicator (a spline indicator variable *FLIB*), a monetary regime shift dummy (for the 1983(2)–85(4) period, denoted by N) interacting with *RPRIME* and its difference, and finally an indicator dummy variable for the 1991(3)–92(2) drought (*DUM92*).

In order to assess the forecasting power of this equation, we need to develop alternative techniques. This is what we turn to in the next section. But it should be noted that several dummy variables which enter the equation could only be defined ex-post. Hence, for ex-ante forecast comparisons, these should be omitted from the equation (at least *DUM92*, and potentially N and *FLIB*, although these could be kept since they were 'predictable' from the government's decisions). Finally for an ex ante forecasting exercise from a past origin, there remains the issue of the I(2) trend: indeed, for computational ease, we do not estimate the state-space model for each subsample but store the whole-sample trend as it was estimated by STAMP and use it as a regressor, assuming, as seems likely, that its value at some date $t < T$ from an estimation over 0,..., T does not strongly differ from the value obtained at t from estimation over $0, ..., t$. However, in order to avoid this issue, we also replace it, in an alternative model, with a deterministic trend.

Thus, the AM technique and its variants are not seen as a well-specified model of the process generating the South African GDP, but rather as a misspecified DMS forecasting tool.

2 Competing forecasts

The forecasting equation developed by Aron and Muellbauer includes a few exogenous variables which ought to be modelled for a proper IMS technique to be used. Unfortunately, doing so would increase the degree of mis-specification, which would, in turn, be detrimental to our assessment of forecast accuracy. Indeed, the aim of this paper is to analyze the *causes* of improved forecast accuracy, and not a mere *observation* of its occurrence. Hence, we resort to two simpler multivariate models which, once solved out, should provide the possibility for both IMS and DMS forecasting. The first model follows a small monetary system for South Africa, the second uses the main variables of the AM equation.

Another issue arises regarding the variable to be forecast: the AM technique provides forecasts of $\Delta_4 y_{T+4}$ from an end of sample T. This annual difference does not fit the definition of the IMS forecast as usually referred to. Indeed, if a model in differences provides some $\Delta\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{B}}\Delta\mathbf{x}_T$, with $y_t = \mathbf{P}\mathbf{x}_t$, then the iterated forecast is given by:

$$\Delta\widehat{\mathbf{x}}_{T+4} = \widehat{\mathbf{B}}\Delta\widehat{\mathbf{x}}_{T+3} = \widehat{\mathbf{B}}^4\Delta\mathbf{x}_T, \tag{1}$$

and hence

$$\begin{aligned} \Delta_4\widehat{\mathbf{x}}_{T+4} &= \widehat{\mathbf{x}}_{T+4} - \mathbf{x}_T = \Delta\widehat{\mathbf{x}}_{T+4} + \Delta\widehat{\mathbf{x}}_{T+3} + \Delta\widehat{\mathbf{x}}_{T+2} + \Delta\widehat{\mathbf{x}}_{T+1} \\ &= \left(\sum_{i=1}^{4}\widehat{\mathbf{B}}^i\right)\Delta\mathbf{x}_T \equiv \left(\widehat{\mathbf{B}}^{\{5\}} - \mathbf{I}\right)\Delta\mathbf{x}_T, \end{aligned}$$

so that we observe that the iterated estimator can become rather complex. Indeed, if, like in AM, the model is $\Delta\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{A}}\mathbf{x}_T$, then:

$$\Delta\widehat{\mathbf{x}}_{T+2} = \widehat{\mathbf{A}}\widehat{\mathbf{x}}_{T+1} = \widehat{\mathbf{A}}\left(\Delta\widehat{\mathbf{x}}_{T+1} + \mathbf{x}_T\right) = \widehat{\mathbf{A}}\left(\widehat{\mathbf{A}} + \mathbf{I}\right)\mathbf{x}_T. \tag{2}$$

Thus, as this very choice of target seems to benefit direct multi-step estimation, we assume that the forecasts to be evaluated concern the levels of the GDP and not the annual differences. Hence from the AM equation $\Delta_4 \widehat{y}_{T+4} = \widehat{c}_1 y_T + \widehat{\mathbf{C}}\mathbf{x}_T$, we retrieve $\widehat{y}_{T+4} = (1 + \widehat{c}_1)\, y_T + \widehat{\mathbf{C}}\mathbf{x}_T$.

2.1 A small monetary model

2.1.1 Data description

The variables which we include in our model comprise the M1 narrow money aggregate (denoted by M), the consumer price index (CPI), the 3–month treasury bill interest rate (per annum, R) and the South African Rand/US dollar exchange rate which were obtained from the International Financial Statistics database provided by the International Monetary Fund. Unfortunately, the IFS series for the real Gross Domestic Product do not seem reliable and hence were discarded and were re-created from the data provided by Aron and Muellbauer (from their series in difference).

Following Jonsson (1999) we conduct a cointegration analysis of five variables:[1,2] the log of real narrow money $m - cpi$, the log of real exchange rate rer, the log of real GDP y, the nominal treasury interest rate R and inflation Δcpi. Augmented Dickey-Fuller tests with 4 lags do not reject the hypothesis that these variables are all integrated of order 1. Figure 1 presents graphs of the nominal and real money together with the inflation rate and their differences; interest rates, the GDP and the real exchange rate are recorded on figure 2. Visual inspection thus confirms the tests.

We notice also the large contractions of real narrow money in the late 1970s and mid 1980s during periods of higher inflation and the continuous depreciation of the Rand with respect to the US dollar.

In order to see how the variables interact, we first report in table 2 the correlation matrix of the economic variables mentioned above. We notice that they are all positive for the submatrix excluding the real exchange rate and that rer (US dollars per rand) is negatively correlated with all the others, being the only consistently decreasing of all. Moreover, the variables which are most correlated are obviously those which are almost monotonic, namely $(m - cpi)$, y and rer, but surprisingly the real money and the inflation are hardly correlated at all.

2.1.2 A VAR system

In addition to the five economic variables, we allow for a constant and a trend to enter the cointegration space. We restrict our attention to a VAR(2) as tests showed that lags beyond 2 are not significant. The VAR in levels, estimated over the 1966(1)–2001(2) period, seems to fit the data

[1] The variables modelled by Jonsson (1999) were the interest rates, real income, exchange rate, broad money and prices.

[2] Computations and tests were conducted using PcGive and GiveWin.

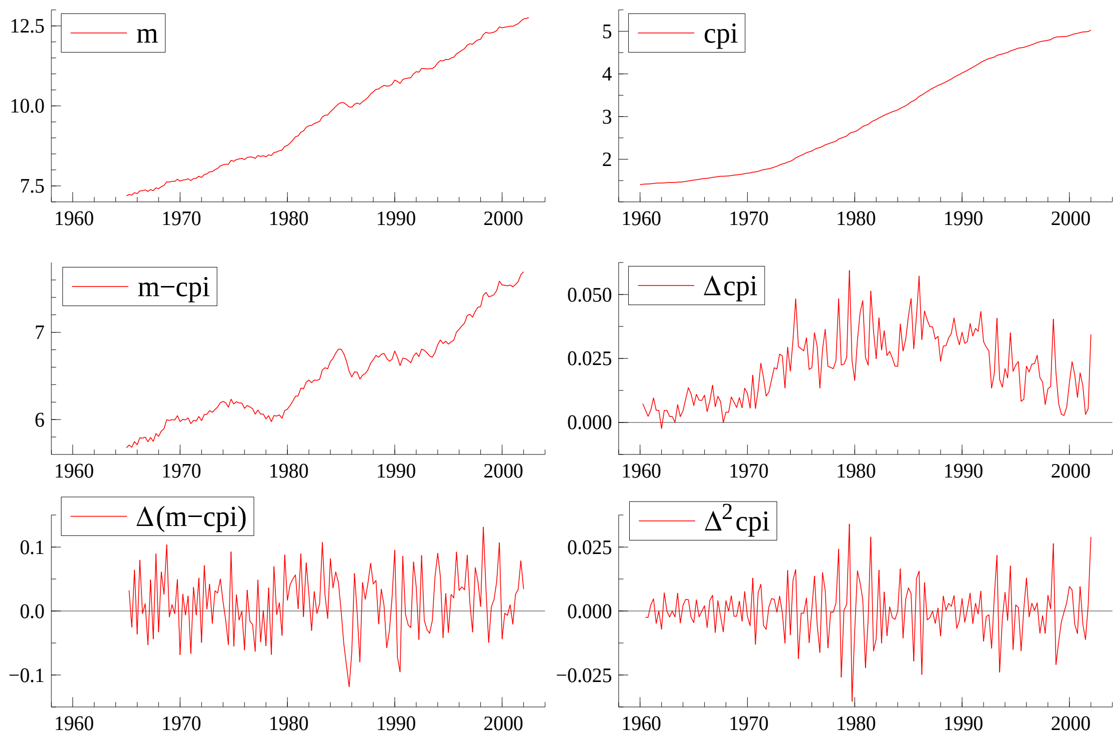

Figure 1: Monetary series and inflation.

Table 2: Correlation matrix of the VAR

Correlations	R	Δcpi	$(m-cpi)$	y	rer
R	1	–	–	–	–
Δcpi	0.366	1	–	–	–
$(m-cpi)$	0.756	0.098	1	–	–
y	0.815	0.364	0.915	1	–
rer	-0.816	-0.201	-0.933	-0.927	1

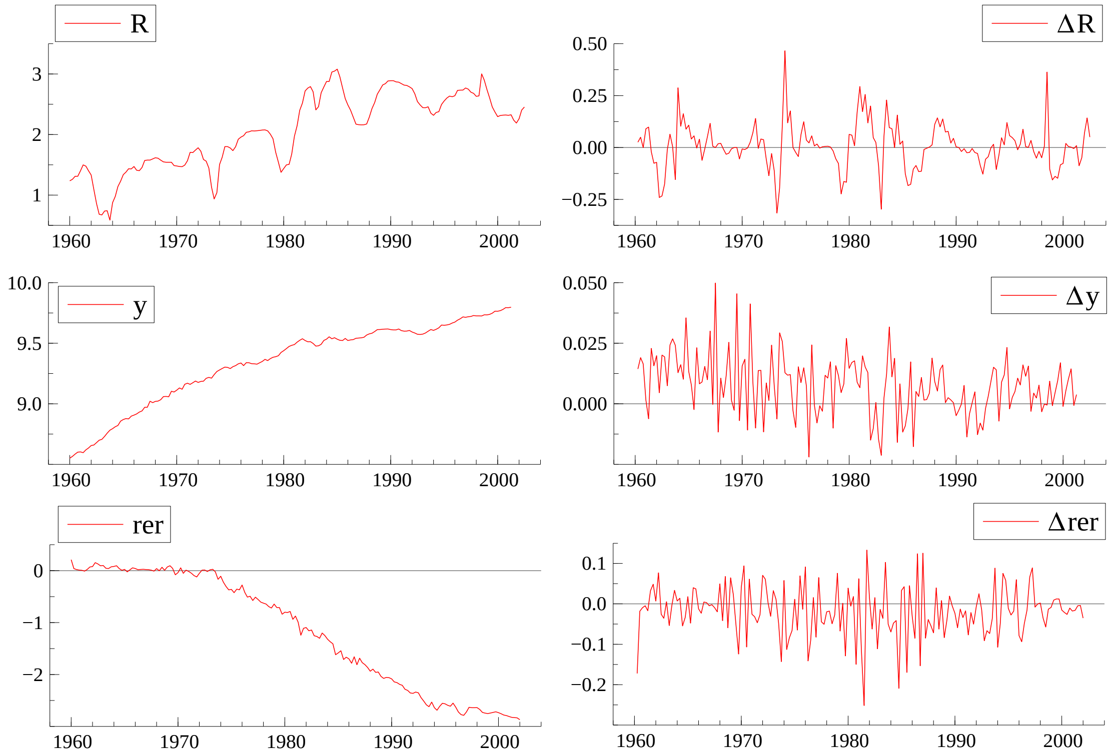

Figure 2: South African interest rates, GDP and real exchange rates.

reasonably well as shown in figure 3. Indeed, except for inflation, the vector of variables is well explained by its past, as can been seen on the figure in spite of the lack of detail.

A corresponding test summary is presented on Table 3 which records statistical information about the VAR, namely the equation residual standard errors ($\widehat{\sigma}$); single-equation evaluation statistics for no serial correlation (F_{ar}, against 5th-order residual autoregression); no ARCH (F_{arch}, against fourth-order); no heteroscedasticity (F_{het}, see White, 1980); and a test for normality (χ^2_{nd}, see Doornik and Hansen, 1994). Analogous system (vector) tests are labelled as ^{v} and, finally, * and ** denote significance at, respectively, the 5 per cent and 1 per cent levels. Hence the Normality tests fail here for most of the variables and the system; this reflects what can be seen from the graphs on the third column of figure 3, namely the presence of large outliers. Given our interest in breaks, and noticing that, but for these, the densities seem close to Normal, we hence retain our model.

Finally, we observe the significant aspect of parameter constancy. Figure 4 presents the equation residuals obtained by recursive estimation and their $0 \pm 2\widehat{\sigma}$ boundaries, which would approximately represent their 95% confidence intervals if the VAR was stationary. Figure 5 records the 1↑ and N↓ (BreakPoint) Chow constancy tests for the individual equations and the VAR system (see

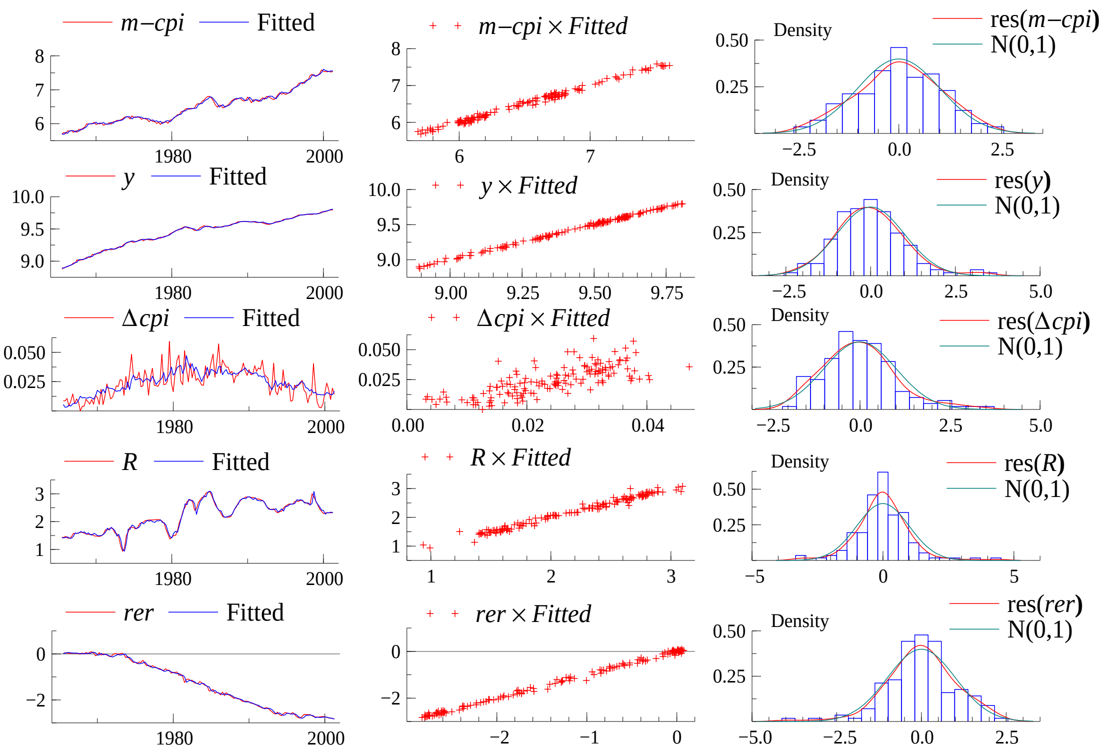

Figure 3: Goodness of fit of the VAR model.

Table 3: VAR Statistics

Statistic	R	Δcpi	$(m - cpi)$	y	rer	VAR
$\widehat{\sigma}$	8.03%	0.86%	4.84%	1.11%	6.29%	
$F_{ar}\,(5, 127)$	0.517	1.22	2.03^{*}	0.108	1.41	–
$F_{arch}\,(4, 124)$	2.39	0.684	0.715	0.740	0.625	–
$F_{het}\,(22, 109)$	1.80	0.773	0.931	1.25	1.09	–
$\chi^2_{nd}\,(2)$	40.6^{**}	10.8^{**}	0.805	7.48^{*}	10.1^{**}	–
$F^{v}_{ar}\,(125, 511)$	–	–	–	–	–	1.06
$F^{v}_{het}\,(330, 1228)$	–	–	–	–	–	1.06
$\chi^{2v}_{nd}\,(10)$	–	–	–	–	–	66.1^{**}

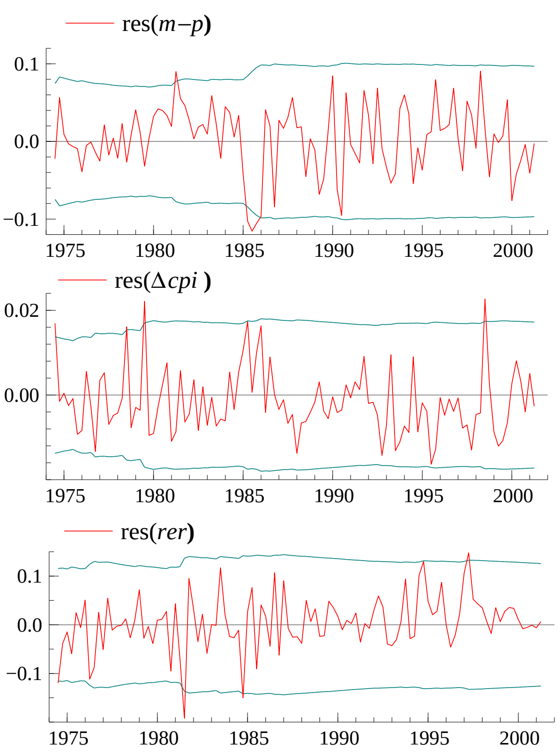

Figure 4: Residuals from recursive estimation of the VAR.

Chow, 1960), together with the 1% boundary. We notice that, according to both figures, we can be reasonably confident in the overall constancy of the relations, although they exhibit occasional breaks and we therefore retain this system for further analysis.

2.1.3 Stationary analysis

Assuming, now, that the VAR is reasonably specified, we would normally investigate the cointegration properties of the five variables. The cointegration statistics reported in table 4—where a constant and a trend entered the cointegration space respectively unrestrictedly and restrictedly—support the hypothesis that there are two cointegrating relations (see Johansen, 1996).

Unfortunately, as shown by Clements and Hendry (1999), the use of cointegrating relationships which experience breaks tends to worsen the accuracy of the forecasts. Given that South Africa provides such an example—indeed, when estimating the model over various subperiods, the trace statistic provides justifications for varying numbers of cointegrating vectors—we decide to exclude the cointegrating vectors from the VAR. Hence the mapping of the VAR in levels to a parsimonious stationary VAR consists simply in differencing. This provides the first of the two IMS forecasting models.

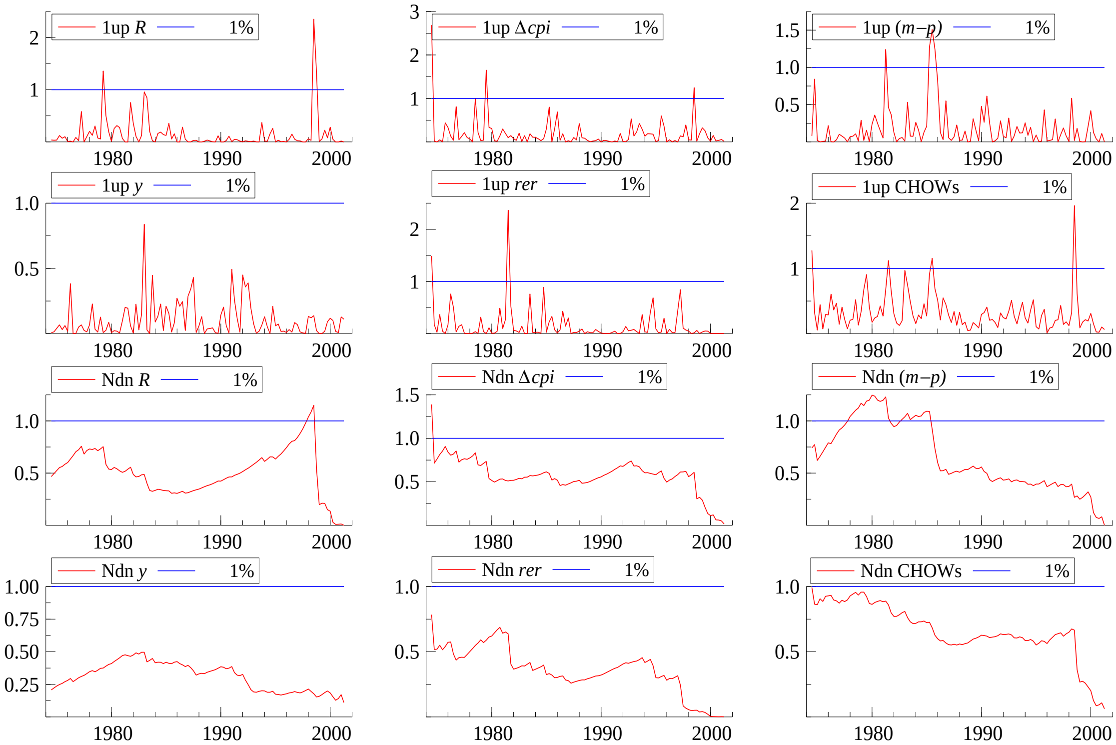

Figure 5: 1↑ and N↓ (BreakPoint) Chow constancy tests for the individual equations and the system.

Table 4: Cointegration statistics: eigenvalues (λ), log-likelihood (l), Trace statistic (Tr) and corresponding pvalue.

rank	λ	l	Tr	*pvalue*
0	–	1457.141	156.95**	[0.000]
1	0.453	1499.968	71.293**	[0.009]
2	0.243	1519.728	31.773	[0.407]
3	0.124	1529.099	13.031	[0.737]
4	0.058	1533.359	4.5108	[0.671]
5	0.031	1535.615	–	–

2.2 An IMS version of AM

In order to analyze the properties of the *multi-step* AM forecasting procedure, we develop below an IMS equivalent. It must be noticed that the parsimonious version of the AM only uses few variables: besides the GDP itself and dummies, the regressors entering the equilibrium correction mechanism are *RPRIME* and its differences, *RCASUR*, *RGSUR*, the log of *TOT* and an I(2) trend (see definitions in section 1.2 above). Hence it seems natural to develop a VAR model which includes these variables. As regards the *N* and *FLIB* indicators, they could be included or not, according to how acute the forecaster's perception of the economic environment could be assumed to have been at the time. Finally, concerning the I(2) trend, which seems an essential element of the model, several strategies can be envisaged: for simplicity and computational ease when forecasting from a past origin, we resort to either omitting it altogether and replacing it with a deterministic trend, or to storing the whole-sample estimated trend and using it as a regressor.

Augmented Dickey-Fuller tests reject the presence of a unit root in the *RCASUR* and *RGSUR* series, but fail to reject it as regards y, *RPRIME* and *TOT*. Thus a cointegration analysis should include the latter three variables, and a trace test for reduced rank claims the existence of one cointegrating relationship between the three, when a trend is allowed to enter the cointegration space. By contrast, if we allow for *N* and *FLIB* to enter unrestrictedly in the system, then we only marginally reject the hypothesis that the matrix is of full rank and hence that the variables do not cointegrate (unreported *pvalue* of 4%, as given by PcGive from estimation over 1960(3)–2000(2), for a VAR(2) model—the lags beyond 2 being statistically insignificant). This small model allows us to generate IMS—and DMS—forecasts which we can compare to the solved out AM equation.

2.3 Univariate methods

We follow the results from the research by Clements and Hendry and use alternative forecasting techniques. These essentially belong to one of two classes: those of differencing or of intercept correcting. Thus, if we wish to forecast y_{T+4} from a forecast origin T, it has been shown that the models:

$$\begin{aligned} \Delta y_t &= \zeta_{1t}, \quad \mathsf{(DV)} \\ \Delta\Delta y_t &= \zeta_{2t}, \quad \mathsf{(DDV\_IMS)} \\ \Delta_4\Delta_4 y_t &= \zeta_{4t}, \quad \mathsf{(DDV\_DMS)} \end{aligned}$$

—where the ζ_{it} are assumed independent white noises—can exhibit some degrees of robustness to breaks. They respectively lead to forecasts $\widehat{y}_{T+4}$ given by:

$$\mathsf{(DV)} \quad : \quad \widehat{y}_{T+4} = y_T, \tag{3a}$$

$$\mathsf{(DDV\_IMS)} \quad : \quad \widehat{y}_{T+4} = y_T + 4\left(y_T - y_{T-1}\right), \tag{3b}$$

$$\mathsf{(DDV\_DMS)} \quad : \quad \widehat{y}_{T+4} = y_T + \left(y_T - y_{T-4}\right). \tag{3c}$$

By contrast, intercept correcting constitutes an adjustment to an existing model, such that if $\widehat{y}_{T+4}$ and $\widetilde{y}_{T+4}$ are forecasts from the IMS and DMS models:

$$\begin{aligned} \mathsf{IMS} &\quad : \quad \widehat{y}_{T+4} = \widehat{\Psi}^4 y_T, \\ \mathsf{DMS} &\quad : \quad \widetilde{y}_{T+4} = \widetilde{\Psi}_4 y_T, \end{aligned}$$

then the intercept corrected forecasts become:

$$\begin{aligned} \mathsf{IMSIC} &\quad : \quad \widehat{y}_{T+4} = \widehat{\Psi}^4 y_T + \left(y_T - \widehat{\Psi}^4 y_{T-4}\right), \\ \mathsf{DMSIC} &\quad : \quad \widetilde{y}_{T+4} = \widetilde{\Psi}_4 y_T + \left(y_T - \widetilde{\Psi}_4 y_{T-4}\right). \end{aligned}$$

Notice that IMSIC could be defined as $\widehat{\Psi}^4 y_T + \widehat{\Psi}^4\left(y_T - \widehat{\Psi} y_{T-1}\right)$, or variations thereof, but these are very unlikely to improve the accuracy and are, hence, left aside.

3 Forecast comparison

3.1 Techniques

We proceed to a comparison of ex-ante forecast accuracy as resulting from the various methods delineated above. These amount to 31 techniques and are labelled as follows: the three 'difference' operator forecasts are DV, DDV_IMS and DDV_DMS as given by (3a), (3b) and (3c). We, then, use six models derived from the AM framework, these are the original AM as defined previously, AM_trend where the stochastic I(2) trend is replaced with a deterministic one, AM_noDUM92 where the 1992 drought is not accounted for, AM_noDUM92_trend which combines both features from the previous two techniques; and finally, the last two techniques ignore *FLIB*, *DUM92* and N altogether: AM_nodum and AM_nodum_trend (dealing as before with the stochastic trend). In order to assess the forecasting power of the small monetary model which we briefly analyzed above, three additional techniques were used to produce forecasts using IMS, DMS and IMSIC procedures, we refer to them as M1_IMS, M1_DMS and M1_IC, and to their equivalent from estimating the model in difference as DM1_IMS and DM1_IC—see (1)—noticing that DMS was not computed here, as it would have involved estimating and combining models at all horizons between 1 and 4. Moreover, the VAR model derived from AM was also used, with the dummy variables in levels and with or without them in differences, thus yielding the seven forecasts labelled as VARAM_IMS, VARAM_DMS, VARAM_IC, DVARAM_IMS, DVARAM_IC, DVARAM_IMS_nodum and DVARAM_IC_nodum. We also computed forecasts from the M1 and VARAM suites, estimated in levels, adding a deterministic trend as a regressor(suffix _trend). Finally, two pooled forecasts from the 6 and 15 overall most accurate methods were also studied (Pool 6 and Pool 15).

3.2 Forecast accuracy

In order to evaluate the competing forecasting techniques, we present the empirical mean-square forecast errors and derive modified Diebold–Mariano test statistics (see Diebold and Mariano, 1995, and Harvey, Leybourne, and Newbold, 1997 who allow for errors to follow a moving average, as in multi-step forecasting) for testing the equality of the MSFEs over various subsamples of the data.

First, as regards the overall forecast accuracy, we notice on table 5 that the AM method as derived by Aron and Muellbauer is the most accurate as it yields the—statistically significantly—lowest MSFE. An interesting feature for direct multi-step estimation lies in the overall accuracy of the variants of the AM technique: indeed, that with the lowest ranking in the forecasting exercise is still more accurate (at the 1% level, see the appendix for the statistics) than the best element from any other class of techniques, even when we remove the dummies from the model and replace the stochastic trend with a deterministic. The next best forecast is obtained by using the small monetary VAR in differences and in levels. This is reassuring since it shows that the purely statistical techniques (DV and DDV) are dominated, although they rank next and before the VARAM model in levels or in differences when we keep the dummies. We notice also that, as regards VAR forecasts (VARAM and M1), the models in differences perform better than the levels; this is in line with the results by Clements and Hendry. In terms of iterated versus direct forecasting, we notice that, when the technique is very inaccurate, DMS is less biased than IMS (see VARAM), but the converse is true when the accuracy of either seems reasonable (see M1). As for IC, it tends to be close to—yet worse than—IMS.

Table 5 also provides the equivalent statistics as computed over smaller samples. Following table 1, we split the dataset into two periods which exhibit different features: in the 1973(4)–

1986(4) subset, the South African economy went through many changes and the breaks were rather frequent, whereas from 1986 to 1995, the country was not very involved in the international economy and the system (i.e. the legal-political environment) did not evolve as fast (which is not to say that the economy did not suffer, say from the 1992 drought). From 1995 to 1998, after the democratic elections, the economic environment (banking and financial sectors) was relatively stable and deregulation took place again afterwards. We, thus, find that forecast accuracy from the econometric models (M1 and VARAM) is improved in the second period, whereas the statistical techniques tend to perform better in the first. Finally, we find that M1 is more accurate in differences in the whole sample and in the first–less stable–period but that the levels are more accurate in the last era (and can even rival some variants of AM) .

The conclusion that can be drawn from the analysis of table 5 is therefore that direct multi-step estimation computed from a dedicated model such as that derived by Aron and Muellbauer is by far the most accurate technique for this dataset. Econometric models come second followed by statistical ones. DMS estimation of a mis-specified VAR (M1, here) does not necessarily improve accuracy. Finally the VAR model which uses the same variables as AM did not perform so well for two reasons: first, the lag length is necessarily reduced by multivariate estimation, and then we did not include a deterministic trend in the VAR in levels. Doing otherwise would have improved accuracy as we show in table 6, but it would not have altered our conclusions.

Indeed, table 6 presents the empirical MSFEs for the same samples as above when a deterministic trend is also used as a regressor, thus providing eight new forecasts. We notice that, as regards the overall accuracy, although the VARAM variants now perform better than beforehand—whether in levels or differences—they yet do not match AM and its derived techniques. In fact, adding a trend improve the accuracy of all models in the first subsample, but in the second it deteriorates that of M1, which was almost on a par with AM. Hence the trend renders the multi-step forecasts more robust in the unstable era. This matches the results from chapter 4 in Chevillon (2004a).

Now, we present on table 7 the empirical MSFEs of the pooled forecasts of the best 15 and 6 estimators. The pooled estimate of the AMs, DM1s, and M1s with and without a trend, which is denoted by Pool 15 performs worse than any of the AMs (except for AM_nodum in the last subperiod) but is more accurate than all the others. Moreover, its empirical MSFE is close in both subperiods. If we restrict pooling to the 6 AMs, then the average forecast ranks third overall and in both subsamples.

3.3 Time series of MSFEs

We analyze, in this subsection, the times series of squared forecast errors. For readability, we do not present all the actual series, but—on figures 6, 7*a* and 8—four-year moving averages thereof. Figure

Table 5: Empirical MSFEs (×10,000) and their accuracy ranking. Stars indicate whether the MSFE is significantly different from the next—in the ranking order—according to the modified Diebold–Mariano test statistic, respectively at the 10% (*), the 5% (**) or the 1% level (***).

	Period								
Forecast	73(4)–00(2)	ranking		73(4)–86(4)	ranking		86(1)–00(2)	ranking	
DV	11.09	14		15.08	12		6.70	12	
DDV_DMS	12.56	15		19.53	17		5.61	9	
DDV_IMS	21.68	18		32.70	19	***	11.42	17	
AM	0.12	1	**	0.17	1		0.07	1	*
AM_trend	1.60	3		0.96	3		2.09	3	
AM_noDUM92	0.37	2	***	0.22	2	*	0.53	2	*
AM_noDUM92_trend	1.68	4	*	1.21	4		2.21	4	
AM_nodum	2.62	6	***	1.44	6	**	3.77	7	
AM_nodum_trend	1.85	5		1.37	5		2.38	5	
M1_IMS	9.03	10		14.82	11		3.60	6	
M1_DMS	10.91	13		15.35	14		7.24	14	
M1_IC	9.47	11		15.34	13		3.85	8	
DM1_IMS	8.52	8		11.45	9		6.19	11	
DM1_IC	8.35	7		10.75	8		5.82	10	
VARAM_IMS	168.13	20		323.15	20		29.49	19	
VARAM_DMS	21.76	19		27.32	18		29.49	20	
VARAM_IC	169.90	21		325.55	21		30.96	21	
DVARAM_IMS	13.32	16		18.13	15		10.44	16	
DVARAM_IC	14.10	17	*	19.01	16		11.49	18	*
DVARAM_IMS_nodum	9.63	12		12.44	10		8.11	15	*
DVARAM_IC_nodum	8.72	9		10.44	7		7.19	13	

Table 6: Empirical MSFEs (×10,000) of the VAR models when a deterministic trend is included in the model.

	Period		
Forecast	1973(4)–2000(2)	1973(4)–1986(4)	1986(1)–2000(2)
VARAM_IMS_trend	11.04	13.52	8.46
VARAM_DMS_trend	13.11	10.31	15.77
VARAM_IMSIC_trend	10.88	14.90	7.10
VARAM_DMSIC_trend	10.65	11.60	10.51
M1_IMS_trend	8.38	8.50	8.56
M1_DMS_trend	11.44	9.62	13.53
M1_IMSIC_trend	10.50	16.58	4.40
M1_DMSIC_trend	7.63	9.14	6.20

Table 7: Empirical MSFEs ($\times 10,000$) of the Pooled forecast of respectively 15 estimators (the AMs, DM1s, and M1s with and without a trend and 6 estimators (the AMs).

	Period		
Forecast	1973(4)–2001(2)	1973(4)–1986(4)	1986(1)–2000(2)
Pool 15	2.78	3.33	4.35
Pool 6	0.85	0.61	1.09

6 records series for some of the most significant techniques from table 5 and we notice that—apart from the AM methods and DVARAM_IMS_nodum—they all see their average forecasting power increase from the 1980s to the 1990s as the MSFEs ($\times 10,000$) hardly venture above ten from 1989 onwards.

Panel 7–*a* records the series for some variants of the AM technique. Comparing, first, AM and AM_trend, we notice that including a deterministic trend does always lead to some loss of forecasting power compared to an I(2), and especially so towards the end of the sample, where we notice on fig. 6 that some of the other models outperform AM_trend. Panel 7–*b* presents the actual series of the squared forecast errors, and we notice that the main benefit from including a stochastic I(2) trend is that it smoothens the large occasional poor forecasts. It should be remembered that this very feature might be simply due to the estimation technique we resorted to.

We report on fig. 8 the other techniques and arrange them according to their historical forecasting power. First, panel *a* presents the only techniques which do not fare too badly in the 1986–1988 interval. For these three procedures—DV, DM1_IC and DVARAM_IMS_nodum—the main loss of forecasting power occurs in the first half of the 1980s, whereas the other techniques tend to underperform essentially over the mid-1980s. Interestingly, despite so many changes in South Africa in the early 1980s, when estimating the levels of the monetary VAR model in solved-out univariate form using PcGets for automated model selection with outlier correction (see Hendry and Krolzig, 2001), a unique dummy variable is picked up by the software in 1983(4)—the beginning of financial liberalization, as accounted for by Aron and Muellbauer via *FLIB*, occurs in 1984(1) (see fig. 9). This corresponds to the sharp increase in inaccuracy in panels *b*–*d*. As a matter of fact, the monetary shift dummy N chosen by AM also starts the transition in 1983(2). It is reassuring to notice that, with such breaks in the levels of M1, the unaffected forecasting techniques are mainly DV and DM1_IC; the success of DVARAM_IMS_nodum should be compared to the failure of DVARAM_IMS although N and *FLIB* correct for the regime transition. The usefulness of the dummy variables appear doubtful in this context.

The next increase in *FLIB* occurs in 1988(1) and lasts until 1990(4); the corresponding changes in the legal-economic environment tend to witness an improvement in forecast accuracy, except—

potentially—DVARAM_IMS_nodum. The end of this period coincides with the 1992 drought, after which three models suffer: DVARAM_IC_nodum, DM1_IMS and DVARAM_IMS. The last two models are typically not designed for robustness to breaks and it is interesting to compare the first to DVARAM_IC which, by contrast, performs quite well over this period. Thus, whereas the use of *FLIB* as a regressor did not improve the forecasts much over 1988-92, *DUM92* proves very relevant. By comparing the techniques whose performance worsens post 1992 to the others, we are led to concluding, by referring to the work by Clements and Hendry, that the 1992 break does imply a step shift for the models in differences.

Finally, post 1997, whereas the AM models tend to lose in accuracy, all the others see their performance improve. This may correspond to the last bout of liberalization, as captured by *FLIB* from 1995(1) to 1996(4).

For comparison of the forecasting techniques with the AM class, we notice that the late 1980s are also—when not accounted for by dummy variables—detrimental to AM, but to a lesser extent. Direct multi-step estimation—in the form developed by Aron and Muellbauer—seems, in conclusion, to provide a very useful method for forecasting macro-economic series in economies which are subject to frequent breaks and regime changes and which are very sensitive to their international environment.

4 Conclusions

The purpose of our analysis was to observe, in an empirical exercise, whether iterated multi-step estimation can improve forecast accuracy and when it does so. We already knew that the conditions most beneficial to DMS are those of model mis-specification and non-constancy of the DGP. Given that Aron and Muellbauer (2002) have derived an equation for forecasting the South African GDP which uses direct multi-step estimation, and that the national economy of this country has undergone several regime changes and extraneous shocks, we decided to build our forecast accuracy comparison on this research.

The strategy for which we settled was to derive several models, and variants thereof, and to record measures of their corresponding historical ex-ante forecast accuracy.

The results are, essentially, that the direct—misspecified—equation derived by AM has impressive forecasting power, whether in troubled periods or more quiet eras. As regards DMS forecasting from alternative models in levels, the conclusions are more mixed: it is hard to recommend one of IMS, DMS or IC as a rule, as their respective performance rankings alter. However, although direct estimation tends not to lead to be best forecast, it does not generally provide the worst either; and it presents the interesting feature of 'decent' accuracy almost everywhere (which is not the case for the other two as VARAM witnesses on table 5).

Direct multi-step estimation is therefore a technique which has the interesting property of being occasionally very close, rarely very far from its target; and its accuracy is increasing in the unpredictability of the economic variable (via regime and deterministic instability).

References

Aron, J. and J. Muellbauer (2002). Interest rate effects on output: evidence from a GDP forecasting model for South Africa. *IMF Staff Papers* **49**, 185–213.

Chevillon, G. (2004a). Multi-step estimation for forecasting economic processes. D.Phil. thesis, Economics department, University of Oxford.

Chevillon, G. (2004b). 'Weak' trends for inference and forecasting in finite samples. mimeo, Economics department, University of Oxford.

Chevillon, G. and D. F. Hendry (2004). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting* **forthcoming**.

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

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-Stationary Economic Time Series.* Cambridge, MA: The MIT Press.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**, 253–263.

Doornik, J. A. and H. Hansen (1994). An omnibus test for univariate and multivariate normality. Discussion paper, Nuffield College.

Harvey, A. C. (1993). *Time Series Models* (2nd ed.). Hemel Hempstead: Harvester Wheatsheaf. First edition: 1981.

Harvey, A. C. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics* **8**, 231–247.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting* **13**, 281–291.

Hendry, D. F. and H.-M. Krolzig (2001). *Automatic Econometric Model Selection.* London: Timberlake Consultants Press.

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector auto-regressive models.* Oxford: Oxford University Press.

Jonsson, G. (1999). Inflation, money demand and purchasing power parity in south africa. Working paper 122, International Monetary Fund.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2000). *STAMP: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultant Press.

Marcellino, M., J. H. Stock, and M. W. Watson (2004). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Mimeo, Harvard University.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838.

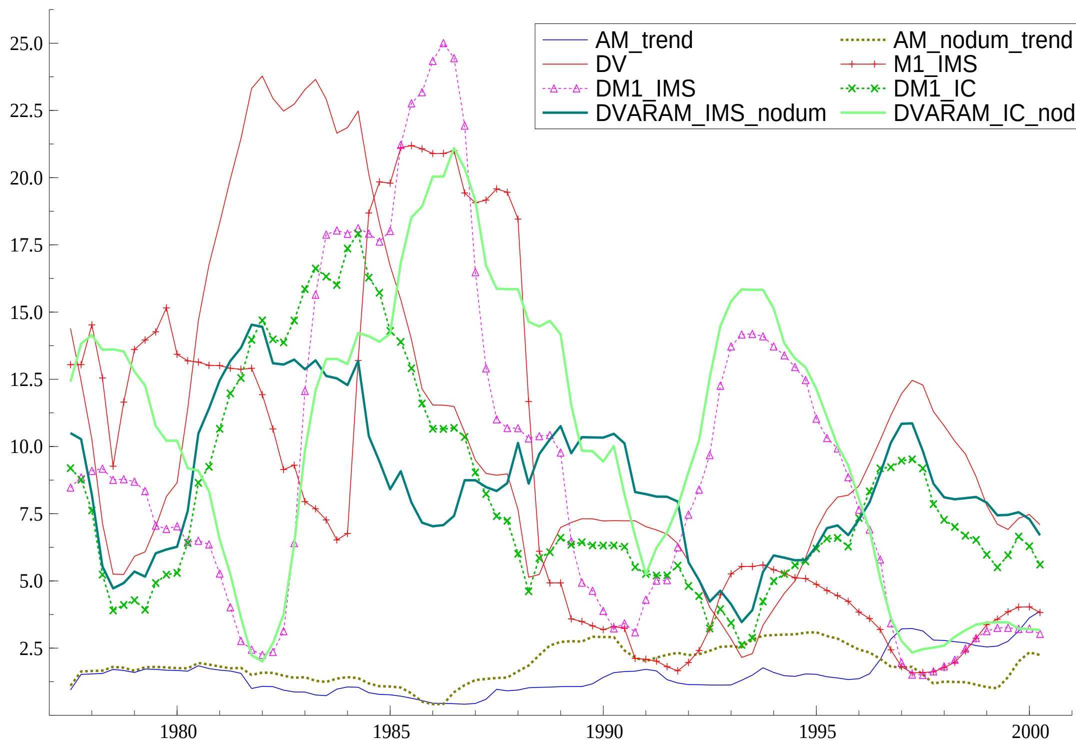

Figure 6: Four year moving averages of the series of Empirical MSFEs ($\times 10,000$).

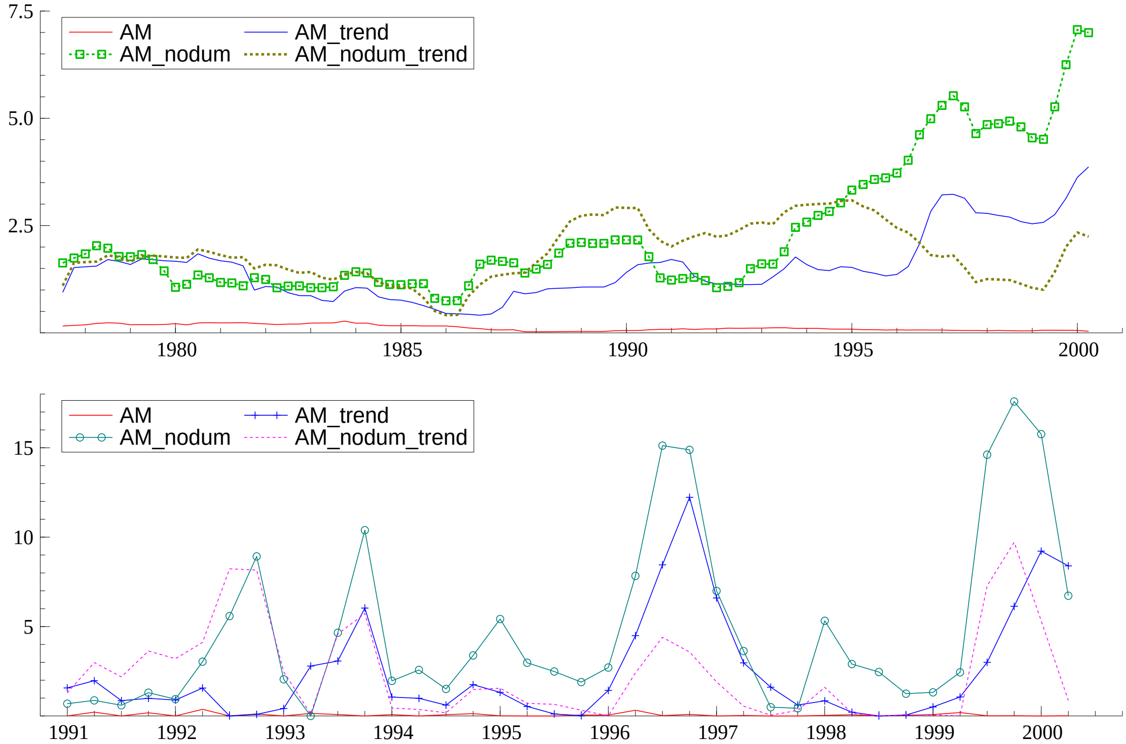

Figure 7: Empirical MSFEs ($\times 10,000$) for variants of the AM model. Panel a records the four year moving average and pannel b the actual series over a subset of the data.

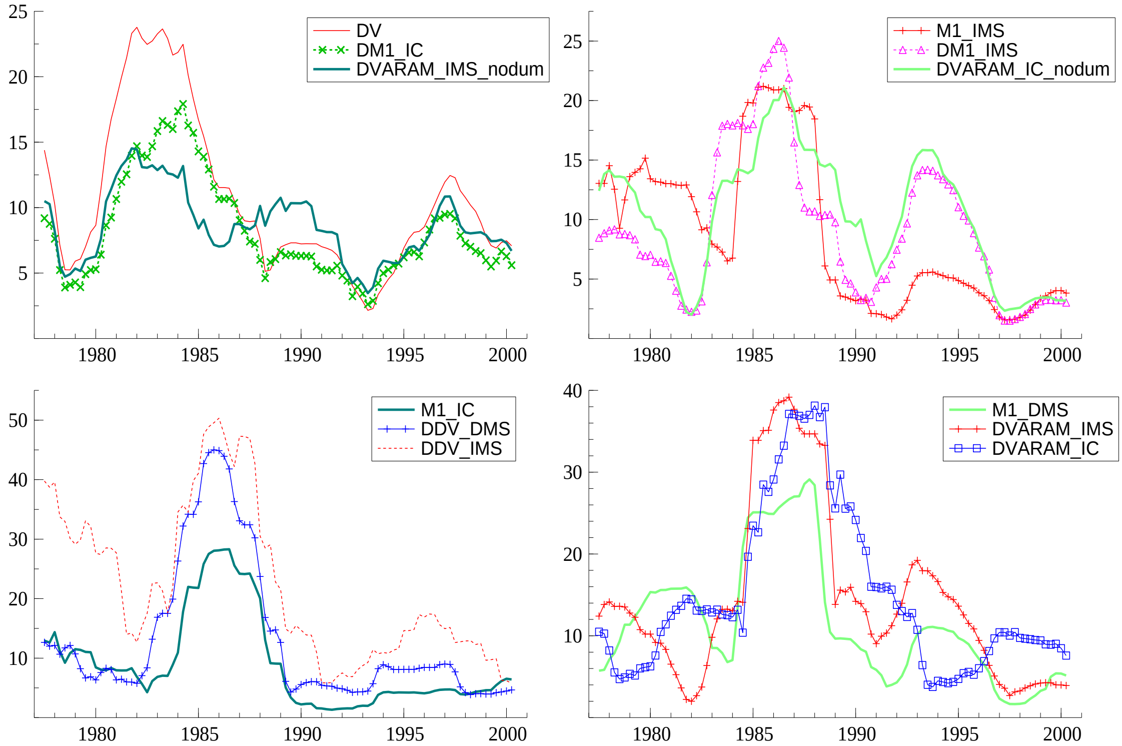

Figure 8: Decomposition of the patterns in the four year moving averages of the series of Empirical MSFEs ($\times 10,000$).

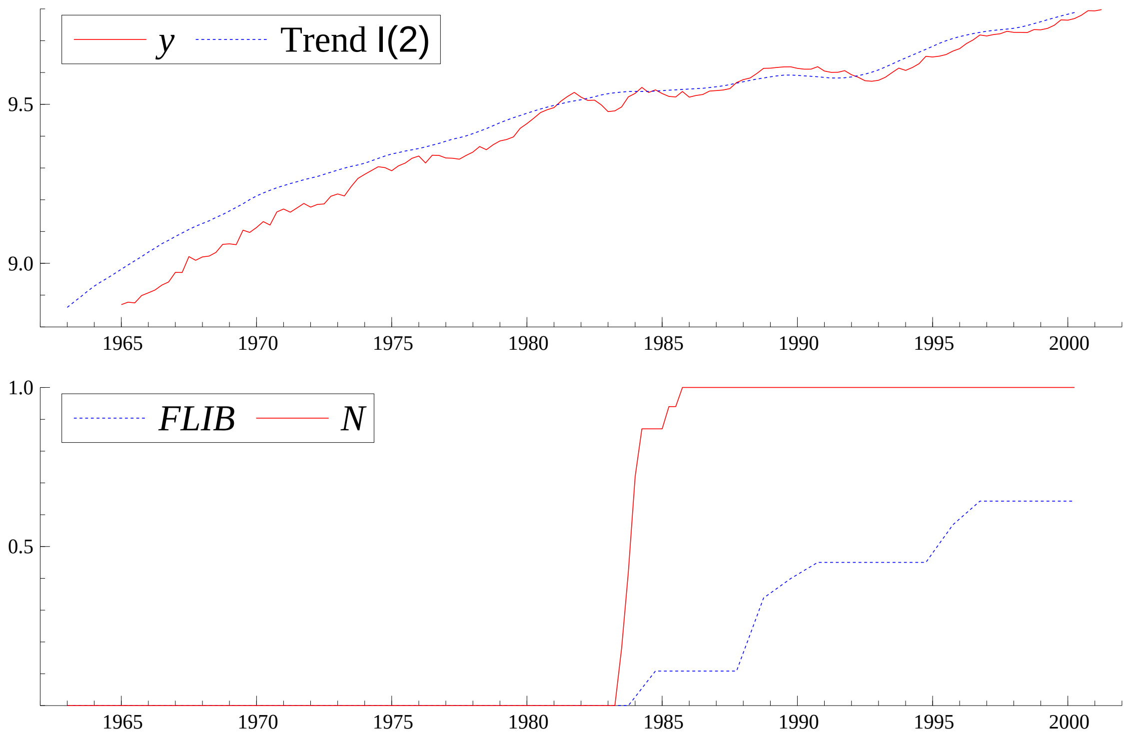

Figure 9: Panel a: Series of the log of GDP and the stochastic $\mathsf{I}(2)$ trend; Panel b: Indicator variables for Financial liberalization ($FLIB$) and the monetary regime transition (N).

Appendix A: Modified Diebold–Mariano Statistics

We record below the statistics of the modified Diebold–Mariano test for equality of the empirical MSFEs as they were computed for the various forecasting techniques used in this paper. The results are split between tables 8 and 9 for length reasons.

Table 8: Modified Diebold–Mariano statistics for testing equality of MSFEs over 1973(4)–2000(2), to be continued on table 9.

Technique	DV	DDV_DMS	DDV_IMS	AM	AM_trend	AM_noDUM92	AM_noDUM92_trend	AM_nodum	AM_nodum_trend	M1_IMS	M1_DMS	M1_IC	DM1_IMS	DM1_IC	VARAM_IMS	VARAM_DMS	VARAM_IC	DVARAM_IMS	DVARAM_IC	DVARAM_IMS_nodum
DV	(–)																			
DDV_DMS	-0.26	(–)																		
DDV_IMS	1.00	2.22	(–)																	
AM	-2.43	-1.38	-2.30	(–)																
AM_detrend	-1.61	-0.84	-1.82	4.91	(–)															
AM_noDUM92	-2.20	-1.25	-2.18	3.48	-3.85	(–)														
AM_noDUM92_trend	-1.55	-0.81	-1.80	6.25	0.36	5.97	(–)													
AM_nodum	-0.97	-0.42	-1.42	5.26	3.95	4.75	2.91	(–)												
AM_nodum_trend	-1.49	-0.76	-1.75	6.30	0.75	6.05	1.96	-2.40	(–)											
M1_IMS	-1.04	-0.79	-2.79	1.23	0.49	1.05	0.45	-0.05	0.39	(–)										
M1_DMS	0.14	0.58	-1.27	2.15	1.48	2.01	1.48	0.96	1.41	2.45	(–)									
M1_IC	-0.88	-0.97	-2.91	1.28	0.56	1.09	0.52	0.02	0.46	0.22	-1.69	(–)								
DM1_IMS	-0.13	0.17	-1.21	2.35	1.47	2.16	1.48	0.82	1.41	0.75	-0.28	0.73	(–)							
DM1_IC	-0.65	0.06	-1.24	3.63	2.18	3.19	2.08	1.09	1.97	0.85	-0.43	0.72	-0.12	(–)						
VARAM_IMS	0.17	0.18	0.14	0.21	0.20	0.21	0.20	0.19	0.20	0.19	0.16	0.19	0.17	0.17	(–)					
VARAM_DMS	2.03	2.09	1.56	2.90	2.71	2.88	2.71	2.54	2.69	2.33	2.09	2.32	2.22	2.22	0.00	(–)				
VARAM_IC	0.18	0.19	0.15	0.23	0.21	0.22	0.21	0.20	0.21	0.20	0.18	0.20	0.18	0.18	1.73	0.01	(–)			
DVARAM_IMS	0.78	2.09	-0.28	2.83	2.17	2.70	2.19	1.66	2.14	2.26	1.10	2.66	1.57	1.15	-0.14	-1.90	-0.15	(–)		
DVARAM_IC	1.22	1.48	0.01	2.95	2.35	2.82	2.33	1.87	2.29	2.12	1.12	2.26	1.15	1.53	-0.13	-1.81	-0.15	0.33	(–)	
DVARAM_IMS_nodum	0.34	0.66	-0.76	3.56	2.49	3.39	2.56	1.67	2.48	1.33	0.24	1.28	1.67	0.76	-0.16	-2.20	-0.17	-0.96	-0.77	(–)
DVARAM_IC_nodum	0.32	0.37	-0.86	4.78	3.29	4.37	3.14	1.95	3.03	1.26	-0.01	1.08	0.30	1.52	-0.16	-2.15	-0.17	-0.78	-1.22	-0.29
VARAM_IMS_trend	0.58	0.71	-0.74	2.73	2.07	2.55	1.98	1.47	1.91	1.66	0.32	1.64	0.55	0.97	-0.15	-1.99	-0.17	-0.46	-0.75	0.09
VARAM_DMS_trend	2.28	1.99	0.76	5.24	4.91	5.10	4.62	4.41	4.49	2.70	1.80	2.77	2.08	3.03	-0.10	-1.43	-0.11	1.08	0.94	1.80
VARAM_IMSIC_trend	0.12	0.36	-1.00	1.91	1.33	1.77	1.29	0.85	1.24	1.12	-0.03	1.06	0.21	0.40	-0.16	-2.07	-0.18	-0.71	-0.96	-0.23
VARAM_DMSIC_trend	1.15	1.31	-0.20	4.68	3.74	4.48	3.68	2.85	3.56	2.17	0.92	2.24	1.21	1.88	-0.14	-1.94	-0.15	0.02	-0.26	0.74
M1_IMS_trend	0.55	0.68	-0.57	5.02	3.51	4.96	3.93	2.45	3.85	1.58	0.41	1.38	1.01	1.16	-0.15	-2.12	-0.16	-0.56	-0.70	0.23
M1_DMS_trend	1.55	1.43	0.34	4.61	3.85	4.61	4.14	3.26	4.11	2.39	1.67	2.09	2.03	2.16	-0.12	-1.64	-0.13	0.72	0.41	1.74
M1_IMSIC_trend	-1.27	-0.35	-1.86	1.55	0.78	1.34	0.73	0.20	0.67	0.44	-0.97	0.25	-0.46	-0.74	-0.18	-2.20	-0.20	-1.39	-1.76	-0.95
M1_DMSIC_trend	-0.18	0.17	-1.22	4.05	2.46	3.63	2.43	1.29	2.29	1.10	-0.40	0.94	0.01	0.20	-0.17	-2.23	-0.18	-1.24	-1.36	-0.74

Table 9: Modified Diebold–Mariano statistics for testing equality of MSFEs over 1973(4)–2000(2), continued.

Technique	DVARAM_IC_nodum	VARAM_IMS_trend	VARAM_DMS_trend	VARAM_IMSIC_trend	VARAM_DMSIC_trend	M1_IMS_trend	M1_DMS_trend	M1_IMSIC_trend	M1_DMSIC_trend
DVARAM_IC_nodum	(−)								
VARAM_IMS_trend	0.48	(−)							
VARAM_DMS_trend	2.84	2.42	(−)						
VARAM_IMSIC_trend	-0.03	-0.90	-2.10	(−)					
VARAM_DMSIC_trend	1.39	0.86	-2.03	1.26	(−)				
M1_IMS_trend	0.61	0.03	-1.86	0.39	-0.69	(−)			
M1_DMS_trend	1.94	1.09	-0.48	1.29	0.78	3.02	(−)		
M1_IMSIC_trend	-1.25	-1.37	-2.56	-0.90	-1.92	-1.20	-2.01	(−)	
M1_DMSIC_trend	-0.48	-0.69	-2.59	-0.25	-1.65	-1.02	-2.19	0.77	(−)