IHS Economics Series
Working Paper 160
July 2004

Parameter Instability and Forecasting Performance: A Monte Carlo Study

Costas Anyfantakis
Guglielmo Maria Caporale
Nikitas Pittis

Institut für Höhere Studien
Institute for Advanced Studies
Vienna

Impressum

Author(s):

Costas Anyfantakis, Guglielmo Maria Caporale, Nikitas Pittis

Title:

Parameter Instability and Forecasting Performance: A Monte Carlo Study

ISSN: Unspecified

2004 Institut für Höhere Studien - Institute for Advanced Studies (IHS)

Josefstädter Straße 39, A-1080 Wien

E-Mail: office@ihs.ac.at

Web: www.ihs.ac.at

160

Reihe Ökonomie

Economics Series

Parameter Instability and Forecasting Performance:

A Monte Carlo Study

Costas Anyfantakis, Guglielmo M. Caporale, Nikitas Pittis

160

Reihe Ökonomie

Economics Series

Parameter Instability and Forecasting Performance:

A Monte Carlo Study

Costas Anyfantakis, Guglielmo M. Caporale, Nikitas Pittis

July 2004

Institut für Höhere Studien (IHS), Wien

Institute for Advanced Studies, Vienna

Contact:

Costas Anyfantakis
University of Piraeus
80, Karaoli & Dimitriou St.
185 34 Piraeus, Greece

Guglielmo Maria Caporale
London South Bank University
103 Borough Road
London SE1 OAA, United Kingdom
☎: +44/20/7815 7012
fax +44/20/7815 8226
e-mail: g.m.caporale@lsbu.ac.uk

Nikitas Pittis
University of Piraeus
80, Karaoli & Dimitriou St.
185 34 Piraeus, Greece

Abstract

This paper uses Monte Carlo techniques to assess the loss in terms of forecast accuracy which is incurred when the true DGP exhibits parameter instability which is either overlooked or incorrectly modelled. We find that the loss is considerable when a FCM is estimated instead of the true TVCM, this loss being an increasing function of the degree of persistence and of the variance of the process driving the slope coefficient. A loss is also incurred when a TVCM different from the correct one is specified, the resulting forecasts being even less accurate than those of a FCM. However, the loss can be minimised by selecting a TVCM which, although incorrect, nests the true one, more specifically an AR(1) model with a constant. Finally, there is hardly any loss resulting from using a TVCM when the underlying DGP is characterised by fixed coefficients.

Keywords

Fixed coefficient models, time varying parameter models, forecasting

JEL Classification

G14, G15, C22

Contents

1 Introduction

Forecasting is one of the most common applications of econometric models. It follows the specification and estimation stages in the modelling process, and it depends crucially on the correct specification of the model and on the selection of the appropriate estimation method for model parameters. The aim of the researcher is to specify a model that approximates the true Data Generation Process (DGP) as closely as possible. A specification with fixed coefficients (FCM) is usually adopted, in the hope that the time heterogeneity properties of the underlying DGP are such that the selected model will be given empirical support by the data. [1] However, misspecification tests often suggest parameter instability which is fundamental, in the sense that it cannot be modelled by augmenting the regressor set with dummy variables or other deterministic components. Moreover, often not even the inclusion of other 'relevant' variables in the model specified by the researcher will remove parameter instability, indicating that practically no FCM is an adequate approximation of the underlying DGP. This has led to the introduction of time varying coefficient models (TVCM), which have been widely used in various fields of empirical economics such as money demand functions, capital asset pricing and exchange rate models (for the latter two, see Wolff 1987, 1989 and Schinasi and Swamy 1989).

A common feature of TVCM regression models is that they assume a specific type of coefficient variation which includes some fixed (hyper) parameters. Specifically, the literature has analysed three types of coefficient variation: the Hildreth-Houck model (1978), the Random Walk model (Cooley and Prescott 1976), and the Return to Normality model (Rosenberg 1973). These models may be cast in State Space Form. Both the estimates of the unknown parameters and the time path of the stochastic coefficients are obtained by using the Kalman Filter recursive algorithm, developed in the engineering literature (see Kalman 1960, and Kalman and Bucy 1961). Theoretical representations of State Space Models and of the Kalman Filter can be found in Meinhold and Singpurwalla (1983), Chow (1984), Hamilton (1994) and Moryson (1998).

Although empirical applications of these models are increasingly common, little is known about their forecasting performance under alternative structures of coefficient variation. The preceding discussion suggests that, if the underlying DGP contains time varying coefficients, the researcher is liable to make two types of errors. The first occurs when he estimates a regression with fixed coefficients

[1] A rather elaborate ARIMA model could appear to be stable even in the presence of parameter instability. However, such a model will break down when used for forecasting purposes (see Harvey 1990).

(FCM), even though the true model exhibits time varying parameters. In this case, he erroneously utilises the Least Squares estimator, or some variant of this estimator. The second type of error refers to the case where the researcher has realised that the DGP contains time varying coefficients, but specifies a type of coefficient variation different from that exhibited by the DGP. Therefore, he correctly estimates a TVCM but selects the wrong type of coefficient variation.

This paper evaluates by means of Monte Carlo simulations the loss in terms of forecast accuracy which is incurred when either error is made, thereby enabling us to suggest an empirical strategy for the applied researcher. Clements and Hendry (1998, 1999, 2002) have analysed extensively the possible reasons for forecast failure, reaching the conclusion that unmodelled shifts in deterministic components are the primary cause. In the present study we ask the question how important overlooking parameter time variation might be as a source of forecast failure. A simple bivariate DGP capturing the salient features of the problem under study is assumed. It consists of a bivariate linear regression where the slope coefficient is time varying, and the regressor follows an AR(1) process. Alternative models of slope coefficient variation are considered, including the three types of coefficient variation mentioned above. Under the assumption that a FCM has been wrongly selected despite the presence of coefficient variation in the DGP, we quantify the resulting loss in terms of forecast accuracy by comparing the forecasts of the wrongly specified FCM with those produced by the 'correct' TVCM. Furthermore, we examine the extent to which the loss depends on the particular form of coefficient variation that characterises the true DGP. As already mentioned, the second type of specification error occurs when the researcher has diagnosed parameter instability in his model, but specifies a TVCM with a structure of coefficient variation different from the true one. In such a case, we are interested in examining a) the loss associated with the wrong selection of the type of coefficient variation for each type of coefficient variation exhibited by the DGP, and b) whether there are cases when the specification of a FCM produces better forecasts than those of a TVCM which assumes the wrong type of coefficient variation.

The paper is organised as follows. Section [2] presents Monte Carlo simulations that address the issues highlighted above. A striking result is that a wrongly specified structure of coefficient variation can produce forecasts that are even worse than those of the 'wrong' FCM. On the other hand, if the model selected for the coefficients is general enough to encompass all possible specifications as special cases, the forecasts are comparable to the forecasts of the 'correct' TCVM. Section [3] summarises the main findings from the Monte Carlo analysis and their implications for the applied researcher.

2 Monte Carlo Analysis

In this section, we investigate by means of Monte Carlo simulations the forecasting cost incurred when the true DGP is a TVCM, and the researcher has assumed either a FCM or a TVCM with a structure of coefficient variation different from the one characterising the DGP. In our experiments the DGP consists of a bivariate linear regression where the slope coefficient is time varying:

$$y_t = b_0 + b_t x_t + e_{1t} \tag{1}$$

with

$$b_t = r_0 + r_1 b_{t-1} + e_{2t} \tag{2}$$

We further assume that the regressor, x_t, is a zero mean, persistent, stationary AR(1) process:

$$x_t = \rho x_{t-1} + e_{3t} \tag{3}$$

Finally, we assume that the errors are iid jointly normal random variables, with zero means and covariance matrix Σ :

$$\begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} \sim NIID \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \tag{4}$$

Concerning the slope coefficient variation, (2) nests all the structures analysed in the literature, that is

(i) Zero Mean AR(1): $r_0 = 0$, $r_1 < 1$

(ii) Random Walk Model: $r_0 = 0$, $r_1 = 1$.

(iii) Hildreth-Houck (HH) Model: $r_0 \neq 0$, $r_1 = 0$.

(iv) AR(1) with constant (Return to Normality Model): $r_0 \neq 0$, $r_1 < 1$.

The number of replications and the sample size for each experiment are 1000 and 100 respectively. We also set $b_0 = 0.5$ and $\rho = 0.75$. For each replication we estimate the linear regression model either assuming a fixed coefficient $b_t = b \forall t$ and applying OLS, or assuming a particular structure of coefficient variation (which, of course, might be different from the one generating the data), and obtain maximum likelihood estimates of the model parameters by using the Kalman Filter (see, for example, Harvey and Phillips 1982, or Chow 1984 for a clear exposition of this method)[2]. If the process followed by the stochastic coefficient has a steady state, then its equilibrium mean and variance are chosen as initial conditions for the algorithm. On the other hand, when there is no steady state (e.g. the random walk case) the initial conditions are usually set as follows: $\mu_0 = E(b_{0|0}) = 0$ and $\Sigma_{0|0} = k \cdot I$, with k being an arbitrarily chosen large number. We follow Koopman, Shephard and Doornik's (1999) recommendation to set $k = 10^6$ initially, and then to rescale it by multiplying by the largest diagonal element of the residuals covariances. Having estimated the model over a (pseudo) in-sample estimation period [1,T], we generate forecasts, $\widehat{y}_{T+h}$, at the horizons h=1,2,3,5 and 10, over a (pseudo) out-of-sample period. To obtain these forecasts we either use realised values of x_{T+h} (ex-post forecasts) or generate forecasts $\widehat{x}_{T+h}$ by utilising (3) (pure ex-ante forecasts). The forecasts are evaluated according to the usual Root Mean Square Error (RMSE) criterion. We also include the forecasts of y_{T+h} produced by the naive ARMA(1,1) model for y_t as a natural benchmark. We consider each of the above four cases of coefficient variation separately in the following subsections.

2.1 Zero Mean AR(1): $r_0 = 0,\ r_1 = 0.845$.

For this set of simulations, we set $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$ and therefore $var(b_t) = \frac{1}{1-0.845^2} = 3.5$. The results for this case are reported in Table 1A. It is apparent that, even when utilising the realised values of x_{T+h}, the 'wrong' FCM is outperformed by the 'correct' TVCM (that is, the one assuming that b_t follows a zero mean AR(1) process) in terms of forecast accuracy, as shown by the RMSE. Moreover, its forecasts are less accurate than even those of the naive ARMA(1,1) benchmark model for all values of h.

Let us assume that, somehow, the researcher realises that the DGP exhibits time varying coefficients, but fails to recognise the exact pattern of coefficient variation, thus specifying a wrong form of coefficient variation. For instance, he assumes that b_t follows a random walk instead of selecting the correct model for b_t, which is a zero mean AR(1) process. We find that the RMSE of the 'wrong' TVCM assumed by the researcher is bigger than the RMSE that would have been obtained had the 'correct' TVCM been selected. For h=5 and 10 the RMSE produced by the wrong TVCM is even greater than the RMSE of either the FCM or the naive ARMA(1,1) benchmark. A similar picture emerges when the researcher chooses the HH model instead of the correct zero mean AR(1) specification for b_t.

[2]Alternatively, the Flexible Least Squares method for recursively estimating the time path of the slope coefficient may be used instead (see Kalaba and Tesfatsion 1989, 1990 and Lutkepohl and Herwartz 1996).

In this case the RMSE produced by the HH model is, as expected, bigger than the RMSE of the correct TVCM for all the forecast horizons h, and also bigger than the RMSE of the TVCM model that assumes that b_t follows a random walk for $h = 1, 2$ and 3. That is, choosing the HH specification causes the worst forecast deterioration for $h = 1, 2$ and 3, while the Random Walk specification does so for $h = 5$ and 10. The misspecification effects on forecasting are minimised when the researcher assumes an AR(1) model with constant which encompasses the zero AR(1) process as a special case. This is due to the fact that the erroneous inclusion of a constant in the AR(1) model does not have any major effects on the accuracy of the estimates of r_0 and r_1, which are close to their true values of 0 and 0.845 respectively.

The results discussed above concern a particular case of coefficient variation, when $var(b_t) = 3.5$. Two additional questions need to be answered. First, is the relative RMSE ranking of the above specifications affected by changes in $var(b_t)$? Second, as changes in $var(b_t)$ could be driven by changes either in σ_{22} or in r_1, what is the relative importance of these two possible sources of change (persistence of coefficient variation (r_1) versus variance of coefficient innovation (σ_{22}))? To answer these questions we conduct two additional extensive simulations. In the first, we keep σ_{22} constant and equal to unity, and let r_1 vary such that $var(b_t)$ takes different values in the interval [1.5, 3.5] by steps of 0.5. In the second, we set r_1 equal to 0.845 and allow σ_{22} to vary in such a way that $var(b_t)$ takes values in the interval [0.5, 3.5] by steps of 0.5. The results for the first case, where r_1 varies, are reported in Figures 1A and 1B for forecast horizons $h = 1$ and $h = 5$, respectively. For the sake of brevity, the results for the second case, where σ_{22} varies, are not reported but briefly summarised below. In the first case, where $var(b_t)$ changes due to changes in r_1, we find that for the shortest forecast horizon, i.e. for $h = 1$, the FCM produces the worst RMSE for all values of r_1. It is important to note that the discrepancy between the FCM and the true TVCM becomes bigger, the bigger is the influence of past shocks inducing coefficient variation, i.e. the bigger is r_1. The same holds true when the HH specification is employed. By contrast, when the researcher erroneously utilises the Random Walk specification the discrepancy between the RMSEs of the correct and erroneous specifications decreases with r_1, since the memory properties of the true model are close to those of the specified one. For all values of r_1 the AR(1) with constant specification appears to be the safest choice if the 'correct' form of coefficient variation cannot be established. For $h = 5$, the discrepancy between the RMSE of the FCM and the true TVCM increases as $r_1$becomes bigger, but is smaller in comparison with the one-step ahead forecast. This result suggests that the loss in short-term forecast accuracy is actually higher when the researcher assumes the FCM instead of the correct TVCM one. It is almost the same if either the HH specification or the FCM are selected. As in the FCM case, the loss increases with the variance of the coefficient variation structure. A significant loss is incurred when the researcher specifies an incorrect Random Walk model for b_t,especially for small values of r_1. This is hardly surprising, since in this case it is erroneously assumed that b_t is not mean reverting, whereas in fact b_t reverts to zero at a speed which is inversely related to r_1.

The results are similar for the second case, where $var(b_t)$ changes due to changes in σ_{22}. The discrepancy between the RMSE of the true TVCM and of all the alternative incorrect specifications increases monotically as σ_{22} increases, the worst forecasting performances being exhibited by the FCM and the HH specification for $h = 1$, and by the random walk specification for $h = 5$.

Finally, the ranking of the various models in terms of forecasting performance stays the same when forecast rather than realised values of x_{T+h} are used (these results are not reported for the sake of brevity).

2.2 Hildreth Hook Model: $r_0 = 0.8,\ r_1 = 0.$

In this case, we assume that the slope coefficient b_t evolves according to the HH model, that is $b_t = 0.8 + e_{2t}$. We set $\sigma_{11} = \sigma_{33} = 1$, but we now set $\sigma_{22} = 3.5$, in order to make the $var(b_t)$ for this case equal to that of b_t in the previous AR(1) case, which was $var(b_t) = \frac{1}{1-0.845^2} = 3.5$. We compute forecasts for all the models of b_t discussed above, in order to evaluate the forecast accuracy loss resulting from misspecification in each case. Table 2A reports the results based on using the realised values of x_{T+h}. As can be seen, the loss due to erroneously using a FCM when the DGP is characterised by time varying coefficients of the HH type is less apparent than before for long horizons. For example, for $var(b_t) = 3.5$ the RMSE produced by the FCM for $h = 5$ and $h = 10$ is equal to 2.21 and 2.26 respectively for the previous case of $r_1 = 0.845$. The corresponding values of the RMSE are 2.08 and 2.17 for the HH specification and for the same level of coefficient variation, that is for $var(b_t) = 3.5$. This suggests that persistence in coefficient variation has a 'net effect' on the forecasting performance of the FCM, over and above its effect through $var(b_t) = \frac{1}{1-{r_1}^2}$. Consequently, the discrepancy between the RMSE of the FCM and of the 'correct' TVCM, i.e. the forecast accuracy loss associated with the choice of the incorrect FCM, is much smaller and at some forecast horizons hardly distinguishable from that observed in the previous case (where b_t followed a zero mean AR(1) model). This finding is consistent with the evidence presented in Figure 1A and 1B that the discrepancy in terms of RMSE between the FCM and the 'true' TVCM increases with persistence in coefficient variation, as measured by r_1. Furthermore, unlike in the previous case of an AR(1) slope coefficient, the forecasting performance of the FCM is not worse than that of the naive ARMA(1,1) forecasting rule.

As for the second type of error, i.e. when the researcher assumes the wrong type of coefficient variation for b_t, we note that the resulting loss is higher than in the previous case: a misspecified TVCM produces forecasts that are less accurate than those not only of the 'correct' TVCM, but also of the FCM at all forecast horizons. In other words, in this case there would be a forecast gain if the researcher overlooked parameter instability altogether and selected instead a FCM. In particular, choosing a Random Walk Model for b_t instead of the correct HH structure generates RMSEs which are higher than those of either the FCM or the naive ARMA(1,1) model at all forecast horizons. The loss due to misspecification is also significant when a zero mean AR(1) model is selected. Therefore, in this case the researcher would be better off using a simple univariate ARMA(1,1) model for y_t rather than a mispecified TVCM for the purpose of forecasting. The loss associated with wrongly specifiying the TVCM is prohibitively large if the true model of coefficient variation is the HH rather than a zero mean AR(1) process as in the previous case. The explanation lies in the very different time series behaviour exhibited by the true HH model and the two alternative models assumed by the researcher. More precisely, in the HH model shocks to the coefficients are 'absorbed' within the same time period and the steady state of the process is equal to 0.8. By contrast, in the AR(1) case shocks do not fade away immediately and the steady state of the process is zero. Finally, in the Random Walk model there is infinite persistence and no steady state for the process. However, if the researcher assumes a model for b_t that encompasses the HH structure as a special case, i.e. if he assumes an AR(1) model with a constant, then the mispecification effects on forecasting are minimised - in this way the researcher allows the process b_t to have a non-zero mean equal to $\frac{r_0}{1-r_1}$, i.e. he does not 'force' the process to fluctuate around an incorrect mean. This is true in both the previous and the present case, which suggests that the best option for the researcher is to specify a model for b_t, e.g $b_t = r_0 + r_1 b_{t-1} + e_{2t}$, that nests the true model of b_t, that is $b_t = r_0 + e_{2t}$. In fact the Monte Carlo means of the estimates of r_0 and r_1 (reported in Table 2B) are very close to their true values of 0.8 and 0 respectively, thus explaining the good forecasting performance achieved by this specification. [3]

[3]However, the over-specification comes at a price, namely the average standard errors are

The same simulations were conducted using different values for $\sigma_{22} = var(b_t)$, specifically for σ_{22} taking values in the interval [0.5, 3.5] by steps of 0.5. Figure 2A plots the RMSE produced by all the alternative forms of coefficient variation, and the FCM, for $h = 1$. The discrepancy between the RMSE of the FCM and the 'correct' TVCM is found to increase with σ_{22}, but is much smaller than in the case where the true process is the zero mean AR(1) model with $r_1 = 0.845$. On the other hand, when the researcher makes the second type of error, i.e. when he employs the incorrect TVCM specification, the discrepancy between the RMSE of the 'true' TVCM and of the TVCM with the zero mean AR(1) specification is inversely related to σ_{22}. Instead, when the Random Walk specification is utilised the RMSE increases monotonically with σ_{22} and the forecast cost is very large. For instance, for $var(b_t) = 3.5$ the RMSE produced by the random walk specification is 25.13% higher than that of the correct HH specification. For the same variability of b_t, that is for $var(b_t) = 3.5$, the RMSE produced by the random walk specification is only 6.13 percent higher than that produced by the correct model with a zero mean AR(1) specification for b_t.

As a general result, we find that the highest forecast accuracy loss is incurred when the model specified by the researcher does not nest the true underlying coefficient structure, though the exact size of the loss will depend on both the selected specification and the degree of variability of the time varying coefficients.[4]

2.3 Non-Zero Mean AR(1): $r_0 = 0.7,\ r_1 = 0.30.$

In this case it is assumed that the slope coefficient follows an AR(1) process with a constant term equal to 0.7. We also set $\sigma_{11} = \sigma_{33} = 1$ and $\sigma_{22} = 3.185$, so that $var(b_t) = \frac{3.185}{1-0.3^2} = 3.5$, namely we choose a value for σ_{22} such that the variance of the coefficient process is the same as in the previous cases, although the process is less persistent. Table 3A reports the RMSEs. It can be seen that if the researcher utilises the FCM instead of the 'correct' TVCM the loss in terms of forecast accuracy is only slight. As in the previous two cases and as shown by Figure 1A, the discrepancy between the RMSE of the FCM and of the 'correct' TVCM decreases as the influence of past shocks to the coefficients diminishes, i.e. when r_1 approaches zero. Moreover, both the 'wrong' FCM and the 'correct' TVCM outperform the naive ARMA(1,1) benchmark model according to the

higher.

[4]The picture remains almost unchanged if forecast (as opposed to realised) values of the explanatory variables are used for forecasting. The results (not reported) again suggest that the researcher should choose the FCM or a simple ARMA(1,1) model in order to avoid the effects of misspecification in the structure of b_t.

RMSE criterion.

Failure to specify the correct form of coefficient variation has serious effects on forecast accuracy. In the present case, if the researcher assumes that b_t follows a zero mean AR(1) process, instead of the true AR(1) process with a constant, he obtains forecasts that are worse, for all horizons, than those not only of the 'correct' TVCM but also of the FCM. This is not surprising, since the zero mean AR(1) specification restricts the coefficient to move around zero while in the 'correct' AR(1) specification $E(b_t) = \frac{r_0}{1-r_1} = 1$. The loss increases further if the researcher specifies a Random Walk model for b_t. It is worth noticing that the RMSE of the HH specification, which imposes no memory restrictions on the coefficient, is almost equal to that of the 'correct' specification. This can be explained by considering the mean estimates of the constant term for the misspecified HH model (see Table 3B): the average value of the estimate of the constant across all replications in the HH specification is close to the mean value of one of the 'correct' coefficient models (AR(1) with a constant). As a result of its unbiased estimate of the mean of b_t, the HH specification matches the forecast performance of the true specification, even though it does not nest the true model for b_t. Similar reasons account for the forecast accuracy of the FCM.

2.4 Random Walk: $r_0 = 0$, $r_1 = 1$.

Here we assume that the slope coefficient of the regression follows a Random Walk process, that is $r_0 = 0$ and $r_1 = 1$ in equation (2) with $\sigma_{22} = 1$. The results are reported in Table 4A. It appears that the consequences in terms of forecasting accuracy of wrongly deciding to employ the FCM are even more serious than in the previous cases considered. For example, if the correct form of coefficient variation, i.e. the random walk, is assumed, then an RMSE as small as 1.55 and 2.54 for forecast horizons $h = 1$ and $h = 5$ respectively, is obtained. If, however, parameter variation is overlooked and a FCM is specified, the RMSE for $h = 1$ and $h = 5$ increases to 5.45 and 6.16 respectively. If parameter instability is detected but the correct type of coefficient variation is not identified, then the forecast accuracy loss depends on the selected model for b_t. If the HH model, which specifies a process for b_t with zero persistence, is chosen, then the loss is comparable to that incurred by estimating a FCM. On the other hand, if an AR(1) model is specified for b_t, either with or without a constant, the associated RMSEs are not much higher than those of the correct Random Walk specification. Table 4B reports the average value of the estimates of r_0 and r_1 obtained if the researcher erroneously specifies b_t as an AR(1) model with constant. It can be seen that the mean estimates of r_0 and r_1 are close to zero and one respectively, which implies that the consistency of the estimates of r_0 and r_1 neutralises, to a large extent, the effects of over-specification. Nevertheless, the loss associated with specifying an AR(1) with constant as the model for b_t is bigger when the true model is a random walk compared to the previous cases where b_t had a steady state. For example, if $h = 5$ then using the AR(1) model with a constant results in an increase in the RMSE of 2.03% and 6.35% when the true process is the zero mean AR(1) model with $r_1 = 0.845$ and the Random Walk respectively. An even bigger loss is incurred when $h = 10$. Consequently, it is crucial that an appropriate method for detecting the correct form of coefficient variation should be found, especially for forecasting at long horizons. The AR(1) specification with a constant produces a smaller percentage loss in terms of forecast accuracy compared to the FCM and the HH specification, and therefore should be used when it is difficult to identify the true process for the coefficient. It is also worth noticing that the loss associated with the erroneous use of the FCM depends on the true type of coefficient variation. It should have become apparent by now that the discrepancy between the RMSE of the 'wrong' FCM and of the 'correct' TVCM is an increasing function of the degree of persistence of the process followed by b_t. In other words, it becomes greater as we move from structures with no or small memory restrictions to structures with infinite memory.

Figure 3 reports the RMSE for all alternative forms of coefficient variation using different values for the variance of the random walk innovations, i.e. σ_{22}. It is clear that the difference between the RMSE of both the FCM and the HH specification and of the correct random walk TVCM is increasing as the variance of the innovations increases. On the other hand, all the autoregressive specifications capture well the dynamics of coefficient variation, thus producing quite accurate forecasts.

To summarise the main findings so far:

(i) If the true DGP exhibits parameter time dependence and the researcher fails to detect it and employs a FCM estimated by OLS, he is likely to incur a significant forecast accuracy loss. For a given level of coefficient variation, i.e. for a fixed value of $var(b_t)$, the loss is minimised when the true coefficient process exhibits a zero degree of persistence.

(ii) If the true DGP is characterised by time varying parameters but the researcher specifies a model of coefficient variation different from the true one, the loss is minimised when the selected model for b_t encompasses the true model as a special case. When this condition is not satisfied, two very different cases can occur. In the first, the model specified by the researcher is such that consistent estimates of the mean of b_t (assuming that the mean exists) are obtained, and the loss is small. In the second, the chosen model does not yield consistent estimates of the mean of b_t (probably because it does not exist), and the loss due to assuming the incorrect TVCM is considerable and comparable to that associated with selecting a FCM.

(iii) Based on the above considerations and the coefficient variation types examined so far, we conclude that the safest strategy for the researcher is to specify an AR(1) model with a constant term as his model for coefficient variation. The reason is that this specification nests all the main models of coefficient variation found in the literature.

The next issue to be addressed is the forecasting performance of the AR(1) model with a constant when it does not encompass the true model for b_t. Below, we investigate its performance under the assumption that b_t follows a process that does not belong to the autoregressive family.

2.5 Moving Average

In this case we assume that the slope coefficient follows a moving average process of the form: $b_t = e_{2t} + 0.5e_{2t-1}$, with $var(e_{2t}) = 2.8$, which implies that $var(b_t) = 3.5$, as in all previous cases. The results obtained using the realised values of x_{T+h} are reported in Table 5, and can be summarised as follows:

(i) The lowest RMSEs are produced by specifications of the process driving b_t that assume that this has a steady state, the AR(1) model with zero mean appearing to be the best choice. These models outperform the FCM in terms of forecasting accuracy at all forecast horizons.

(ii) The highest RMSE is produced by the assumption that b_t follows a random walk. This is not surprising, given the fact that the researcher assumes a process for b_t with vastly different statistical properties (no mean reversion) from the true one (stationary).

(iii) In this case the two AR(1) models (with and without constant) are underfitted, since the true model for b_t is an $AR(\infty)$. The HH specification, which does not include any autoregressive terms, is even more so. However, these specifications perform well relative to the random walk model because the mean of b_t (which is equal to zero) is still estimated with some degree of accuracy, despite the underfitting.

(iv) The ARMA(1,1) forecasting rule produces RMSEs that are more or less equal to those of the TVCMs that assume that the process of b_t has a steady state. This means that there is practically no benefit from employing a TVCM rather than an ARMA(1,1) forecasting rule if the chosen model of coefficient variation does not nest the true process followed by b_t. Some minor gains in terms of lower RMSE derive from using an AR(1) model, but only for a forecast horizon $h = 1$.

2.6 Fixed Coefficients: $b_t = 1$ for all t

Finally, we examine the forecasting performance of the alternative TVCMs when the DGP is characterised by fixed coefficients, but the researcher erroneously assumes that they exhibit time dependence. The results are reported in Table 6. It appears that there is almost no loss resulting from using a TVCM, as the RMSE of the 'correct' FCM and of the alternative 'wrong' TVCMs are hardly distinguishable. Also, the simple ARMA(1,1) forecasting rule produces the worst forecasts at all forecast horizons.

2.7 Persistence and Forecasting Performance

The Monte Carlo analysis we have conducted suggests that the forecasting performance of the FCM improves when the true process driving b_t exhibits a zero or low degree of persistence, compared to the case when b_t follows a random walk. One might ask what the reasons are. Given that the forecasts of y_t in the context of the FCM-OLS framework are computed as $\widehat{y}_{T+h} = \widehat{b}_0 + \widehat{b}x_{T+h}$ the first question to answer is what the OLS estimator 'estimates' when b_t is time varying, i.e what exactly $\widehat{b}$ is. Heuristically, if the process followed by b_t is such that y_t is stationary and ergodic, then one would expect $\widehat{b}$ to converge in probability to $E(b_t)$. To investigate this issue, we carry out additional simulations. Specifically, we examine the finite sample properties of the OLS estimator of the slope coefficient in the context of the FCM. Table 7 reports the Monte Carlo means and standard deviation of the OLS estimator of the slope coefficient applied to the FCM. In particular, it presents the simulation results when the true process followed by b_t is: (i) a zero mean AR(1) (Panel A); (ii) the HH (Panel B); (iii) a stationary AR(1) with constant and a Random Walk (Panel C and D respectively); (iv) the FMC, i.e $b_t = b \equiv 1, \forall t$ (Panel E). In brief, we find that:

(i) When b_t follows an AR(1) process, with $E(b_t) = b$ (where b may be zero), or it follows the HH process with $E(b_t) = r_0 \equiv b$, the OLS estimator $\widehat{b}$ converges to b as the sample size increases. Note that the standard deviation of $\widehat{b}$ decreases at approximately the same rate for all the stationary processes of b_t.

(ii) When b_t follows a random walk, however, no convergence is observed. Instead, the standard deviation of $\widehat{b}$ increases with the sample size. This explains why the forecasting performance of a FCM is extremely poor when the true model of coefficient variation is a random walk.

3 Conclusions

Most econometric models are built either for forecasting or for policy simulation and analysis. Even in the latter case, analysing their forecasting properties can still be important for assessing their adequacy as an approximation for the underlying DGP. A variety of reasons have been considered for the poor forecasting record of many models used by practitioners or as a guide to policymakers. Clements and Hendry (1998, 1999, 2002) have developed a taxonomy of forecasting errors, and concluded that shifts in deterministic factors in the out-of-sample forecast period are the main reason for forecast failure. According to their analysis, provided the deterministic components are adequately treated, other factors (including misspecification of the stochastic components) are not likely to have significant effects on forecasting. In the presence of structural breaks, they suggest using differenced or double-differenced series in addition to intercept correction in the presence of structural breaks in order to reduce forecast bias (though at the cost of a higher forecast variance), whilst are rather dismissive of Markov switching or threshold models, which are reported not to significantly outperform linear AR specifications. In this paper we argue that overlooking time dependency in the parameters is in fact a potentially crucial explanation for forecast failure, which can be remedied by using models allowing for coefficient variation. In practice, time invariant parameters are often assumed, or even when parameter instability is taken into account, the type of time heterogeneity which is allowed for is not as complex as that present in the data. [5] By means of Monte Carlo experiments, we quantify the deterioration in model forecasts which occurs when either time variation in the parameters is ignored, or the wrong model of coefficient variation is chosen, and show that the consequences for forecasting can be severe.

In particular, our findings suggest that there is a considerable loss in terms of forecast accuracy when the true process is a TVCM and the researcher assumes a FCM, this loss being an increasing function of the degree of persistence [6] and of

[5]See Caporale and Pittis (2002) for further details. They also discuss the undesirable consequences for statistical testing, and suggest a strategy to isolate the invariants and estimate a stable subsystem conditional on the other subset of variables (if these are superexogenous with respect to the parameters of interest). In this case, standard statistical inference and estimation techniques are valid.

[6]Since, as discussed before, if b_t follows a highly persistent process, its estimator does not

the variance of the process driving the slope coefficient. [7] A loss is also incurred when the true process is a TVCM and the researcher, although detecting parameter instability, specifies a model for coefficient variation different from the true one. Under these circumstances, surprisingly, the forecasts are even less accurate than those produced by a FCM. However, the loss can be minimised by selecting a model for coefficient variation which, although incorrect, nests the true one, in which case the forecasts are comparable to those of the 'correct' TCVM. The forecasting performance is particularly poor when the true process is a random walk, and either the FCM or the HH models are chosen, as neither of them encompasses the random walk as a special case. Further, even if the true DGP is known and the correct random walk specification is adopted, forecasts at long horizons are extremely unreliable if forecast (rather than realised) values of the regressor are used. Finally, there is hardly any loss resulting from using a TVCM when the underlying DGP is characterised by fixed coefficients, the resulting forecasts being almost as accurate as those from a FCM. Consequently, the applied researcher interested in forecasting, and not having full information about the underlying DGP, would be well advised to estimate a time varying coefficient model, and more specifically an AR(1) model with a constant. This empirical strategy will obviously produce the best forecasts if the selected structure of coefficient variation is in fact the true one; it will generate forecasts almost as precise as those which would have been obtained using the true model of coefficient variation if this is different from the chosen one (the reason being the generality of the AR(1) with a constant specification); it will also guarantee a forecasting performance of the model which nearly matches that of a FCM even if the DGP does in fact exhibit fixed coefficients. Finally, using realised instead of forecast values of the regressor is essential for the purpose of long-term forecasting.

converge.

[7]The very poor forecasting performance of monetary models of the exchange rate documented by Meese and Rogoff (1983) might very well reflect the fact that they overlook parameter instability. Caporale and Pittis (2001) provide clear evidence of such instability, and estimate a stable subsystem.

References

[1] Caporale, G.M. and N. Pittis (2001), "Parameter instability, superexogeneity and the monetary model of the exchange rate, Weltwirtschaftliches Archiv, 137, 3, 501-524.

[2] Caporale, G.M. and N. Pittis (2002), "Unit roots versus other types of time heterogeneity, parameter time dependence and superexogeneity, *Journal of Forecasting*, 21, 3, 207-223.

[3] Chow, G.C. (1984), "Random and changing coefficient models, in Z.Griliches and M.D Intriligator (eds.), Handbook of Econometrics, vol. II, 1213-1245, Elsevier Science, Amsterdam.

[4] Clements, M.P. and D.F. Hendry (1998), Forecasting Economic Time Series, Cambridge University Press, Cambridge.

[5] Clements, M.P. and D.F. Hendry (1999), Forecasting Non-Stationary Economic Time Series, MIT Press, Cambridge, MA.

[6] Clements, M.P. and D.F. Hendry (2002), A Companion to Economic Forecasting, Blackwell Publishers, Massachusetts USA.

[7] Cooley, T.F. and E.C. Prescott (1976), "Estimation in the presence of stochastic time variation, *Econometrica*, 44, 167-184.

[8] Hamilton J.D. (1994), Time Series Analysis, Princeton University Press, Princeton.

[9] Harvey, A.C. (1990), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.

[10] Harvey, A.C. and G.D.A. Phillips (1982), "The estimation of regression models with time-varying parameters", in M. Deistler, E. Furst and G. Schwodiaur (eds.), Games, Economic Dynamics and Time Series Analysis,, Wurzburg and Cambridge, Massachusetts, Physica-Verlag.

[11] Hildreth, C. and J.P. Houck (1968), "Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association*, 63, 584-595.

[12] Kalman, R.E. (1960), "A new approach to linear filtering and prediction problems, *Journal of Basic Engineering, Transactions ASME*, Series D 82, 35-45.

[13] Kalman, R.E. and R.S. Bucy (1961), "New results in linear filtering and prediction theory, *Journal of Basic Engineering, Transactions ASME*, Series D 83, 95-108.

[14] Koopman, S.J., N. Shephard and J.A. Doornik (1999), "Statistical algorithms for models in state space form using SsfPack 2.2, *Econometrics Journal*, 2, 113-166.

[15] Meese, R. and K. Rogoff (1983), "Empirical exchange rate models of the seventies: do they fit out of sample?, *Journal of International Economics*, 14, 3-24.

[16] R.J. Meinhold and N.D. Singpurwalla (1983), "Understanding the Kalman Filter, *American Statistician*, 37, 2, 123-127.

[17] Moryson, M. (1998), Testing for Random Walk Coefficients in Regression and State Space Models, Physica-Verlag, Berlin.

[18] Rosenberg, B. (1973), "Random coefficient models: the analysis of a cross section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement*, 2, 399-428.

[19] Schinasi, G. and P.A.V.B. Swamy (1989), "The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change, *Journal of International Money and Finance*, 8, 375-390.

[20] Wolff , C. (1987), "Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models, Journal of Business and Economic Statistics, 5, 1, 87-97.

[21] Wolff, C. (1989), "Exchange rate models and parameter variation: The case of the Dollar-Mark exchange rate", *Journal of Economics*, 50, 1, 55-62.

Authors: Costas Anyfantakis, Guglielmo M. Caporale, Nikitas Pittis

Title: Parameter Instability and Forecasting Performance: A Monte Carlo Study

Reihe Ökonomie / Economics Series 160

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996