Macroeconomic Dynamics, **12**, 2008, 151–171. Printed in the United States of America.
DOI: 10.1017/S1365100507060385

ARTICLES

MODELING MACROECONOMIC SUBAGGREGATES: AN APPLICATION OF NONLINEAR COINTEGRATION

ADUSEI JUMAH AND ROBERT M. KUNST
University of Vienna

Many macroeconometric models depict situations where the shares of the major demand aggregates in output are stable over time. The joint dynamic behavior of the considered demand aggregate and output may thus be approximated by a cointegrated vector autoregression. However, the shares of many demand subaggregates in output are rather mobile and changing over time. In order to simultaneously capture the flexibility of the shares of the subaggregates and the long-run constancy of the share of the total aggregate, we consider trivariate systems of two macroeconomic subaggregates and output with error-correction terms that are nonlinear functions of the original variables. The merits of the models are evaluated by means of several forecasting experiments.

Keywords: National Accounts, Great Ratios, Forecasting, Predictive Accuracy

1. INTRODUCTION

Econometric forecasting basically has to strike a balance between statistical evidence and plausibility. A case in point is the well-known fact that excellent short-run macroeconomic forecasts can be obtained from models that generate quite infeasible economies in the longer run. For example, good one-quarter forecasts for private consumption can be derived from models that imply enormous or negative household saving in the longer run. Similarly, good short-run forecasts for the unemployment rate may be obtained from models that imply a longer-run rate below zero or above 50%.

Conversely, the benefits of using a good model may not be revealed in short-run prediction. Engle and Yoo (1987) have shown that cointegration is able to improve prediction only at larger forecast horizons, even in simulated structures, where the cointegrated model is known to be true [see also Christoffersen and Diebold (1998)]. However, cointegration expresses plausible and well-accepted economic equilibrium conditions.

Address correspondense to: Adusei Jumah, Department of Economics, University of Vienna, BWZ, Bruenner Strasse 72, 1210 Vienna, Austria. e-mail: adusei.jumah@univie.ac.at.

 1365-1005/07 $18.00

If a forecaster aims at longer horizons—developing scenarios for several decades—pure extrapolation of statistically identified coefficient structures often fails. To attain plausibility, long-run equilibrium conditions have to be accounted for, even if they are rejected on statistical grounds for a sample of limited time range. On the basis of this reasoning, the current paper focuses on the longer-run implications of certain error-correction models for macroeconomic scenarios rather than on statistical tests.

The problem at hand is one that concerns longer-run scenarios, even though it has been mostly ignored in the forecasting literature. Demand aggregates such as private consumer expenditure, in short consumption, and gross fixed capital formation, in short investment, are known to keep to a rather stable share of total output as measured by gross domestic product (GDP). In some theoretical models, the properties of long-run constant ratios of consumption to output and of investment to output follow from the assumptions of collective utility maximization and relatively general forms of aggregate production functions (Romer 1996). In macroeconomic works, this long-run constancy of "great ratios" is usually captured by the econometric condition of cointegration that ties demand aggregates to roughly constant shares of output in dynamic equilibrium (Stock and Watson 1988; Kunst and Neusser 1990; King et al. 1991). Although the long-run constancy of the great ratios in actual data has been controversial from its early days (Klein and Kosobud 1961) and its validity has been criticized on theoretical and empirical grounds, it constitutes one of the typical working hypotheses for macroeconometric model building.

Usually, economic forecasters are additionally required to compile predictions for subaggregates of the main demand aggregates, such as consumption of durable goods and investment in machinery. For these subaggregates, no long-run relationships are known or have been found to hold. For example, many OECD economies show trends of declining shares of investment in construction and rising shares of investment in equipment, possibly due to changes in technology. Imposing cointegration or error correction with regard to the subaggregates may result in implausible longer-run scenarios. Alternatively, imposing long-run constancy on the main demand shares in equations for subaggregates implies error correction models with nonlinear cointegration. These will be outlined in Section 2.

For longer-run scenarios, asymptotic properties of such error-correction models are crucial. Simulating the basic nonlinear cointegration models usually leads to a high probability of one of the subaggregates disappearing from the market, while the other subaggregates take over the whole time-constant share. We show how restricting the deterministic drift part of the cointegrating models may lead to a considerable delay of such unwanted long-run features.

As empirical examples for the techniques, we use data for investment components in the United Kingdom. Minor components are aggregated so that total capital formation is split among the two parts of construction investment, residential structures and the remainder. The scenarios serve to highlight the main longer-run features of the models.

Our main finding is that, although unit-root testing is not really conclusive with regard to classifying the data-generating processes (DGP) of the aggregate as well as the subaggregate investment-output ratios, the nonlinear cointegration model is superior as a forecasting device at the projection horizons of interest of up to ten years. We confirm this finding by means of a parametric bootstrap experiment that assumes the nonlinear model as the DGP. Interestingly, the nonlinear cointegration model even dominates when the experiment assumes that the DGP conforms to a linear cointegration model. A very similar impression was obtained from parallel experiments for Austria and France. For brevity, we do not report them here in detail.

Section 3 provides a description of the data, whereas Section 4 presents the main empirical results. Section 5 presents two small prediction evaluation experiments. The first experiment compares a stochastic forecast with the observed data. The second experiment draws simulated processes via a parametric bootstrap and compares stochastic predictions with the simulated data. Section 6 concludes.

2. METHODOLOGY

2.1. Traditional Error-Correction Models

According to economic theory and also to observation, the share of a major demand aggregate—say, consumption or investment—in output is roughly constant in the long run. If a modeler just wished to fulfil the task of developing a joint model for two considered demand aggregates and output, he or she might approximate dynamic behavior by a cointegrated vector autoregression, otherwise known as a vector error-correction model (VECM),

$$\begin{pmatrix} \Delta c_t \\ \Delta e_t \\ \Delta y_t \end{pmatrix} = \mu + \alpha_1(c_{t-1} - y_{t-1} - \delta_1) + \alpha_2(e_{t-1} - y_{t-1} - \delta_2) + \Gamma \begin{pmatrix} \Delta c_{t-1} \\ \Delta e_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \varepsilon_t. \tag{1}$$

In (1), c and e denote the demand aggregates in logarithms and y denotes the output in logarithms. Depending on the loading vectors α_i, $i = 1, 2$, which typically have negative entries at location i and positive entries at location 3 and on Γ, the model is marginally stable in the sense that first differences $(\Delta c, \Delta e, \Delta y)$ form a stationary and ergodic process, whereas (c, e, y) do not. For some applications, particularly for the consumption-output ratio, the third entry of the corresponding α vector is close to zero and y becomes weakly exogenous for that vector. The model is called error-correcting, as the variables $c - y$ and $e - y$ tend to move back to their long-run equilibria δ_1 and δ_2. Nonlinear functions of $c - y$ and $e - y$, such as the original "great" ratios $\exp(c - y)$ and $\exp(e - y)$, will also remain close to their equilibrium. For a good presentation of the linear VECM and its statistical features, see Johansen (1995).

In large macroeconometric models, many demand aggregates are decomposed into subaggregates. Gross fixed investment is disaggregated into investment on construction and on equipment. Private consumer spending may be disaggregated into spending on services, durable goods, and nondurables. Let g denote the logarithm of a main demand aggregate that is decomposed into two subaggregates. Let c and e now denote the logarithms of the two subaggregates. Then there is a theoretical exact adding-up condition, $\exp(g) = \exp(c) + \exp(e)$. Generalizations to a larger number of subaggregates are straightforward.

In some cases, the share of each subaggregate in output will itself be constant in the long run. Then, one may consider trivariate VECMs for the subaggregates and y. System (1) has this form already. The total aggregate g results from c and e and is not included explicitly. Like $c - y$ and $e - y$, $g - y$ will also be stationary.

In many other cases, individual ratios $\exp(c-y)$ and $\exp(e-y)$ are rather mobile and changing through time, while $\exp(g - y)$ remains stable. For example, after World War II, the share of construction in total investment and hence in total output was decreasing for several decades in most of the war-torn European economies. During that phase, perhaps except in the immediate aftermath of the war, the investment-output ratio was roughly constant. Expenditures on machinery simply replaced expenditures on construction. Similarly, the consumption-output ratio has shown a remarkable constancy over the last few decades, while the components of household expenditure were subject to trends that reflected the increasing wealth and also shifts in taste. A relative decrease in expenditures on nondurables reflects the lesser importance of basic goods. Simultaneously, durable goods showed a relative expansion. Later on, an increased demand for luxury services implied a rising share of services in consumer expenditures. In summary, sizeable shifts occur among the subaggregates, while the total aggregate grows in parallel with the general economy.

2.2. An Error-Correction Model with Nonlinear Cointegration

A trivariate variable $X = (c, e, y)'$ consists of two parts of a demand aggregate and gross output. We consider the model

$$\begin{pmatrix} \Delta c_t \\ \Delta e_t \\ \Delta y_t \end{pmatrix} = \mu + \alpha[\ln\{\exp(c_{t-1} - y_{t-1}) + \exp(e_{t-1} - y_{t-1})\} - \delta] + \Gamma \begin{pmatrix} \Delta c_{t-1} \\ \Delta e_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \varepsilon_t, \tag{2}$$

which contains an error-correction term that is a nonlinear function of the level variables. It is convenient to include an explicit target for the logarithmic "great" ratio δ and to separate it from economic growth represented by μ. As in (1), the third element of α may be close to zero, expressing the fact that there is no tendency in overall output to adjust to equilibrium. In that case, y can be regarded as exogenous for the longer-run characteristics α and δ.

The model (2) is a member of a class of nonlinear dynamic models that was analyzed by Escribano and Mira [2002, (EM); see also Escribano (2004)]. These models are characterized by two main features. First, the equilibrium term is a nonlinear function of a linear transform of the original variables; in EM's notation $J(\beta' X)$ and

$$J(w) = J(w_1, w_2) = \alpha[\ln\{\exp(w_1) + \exp(w_2)\} - \delta].$$

This corresponds to (2) for $X = (c, e, y)'$ and

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}.$$

Second, the function $\partial J(w)/\partial w$ follows a Lipschitz condition. This derivative can be represented as

$$\frac{\partial J(w)}{\partial w'} = \alpha[\{1 + \exp(w_2 - w_1)\}^{-1}, \{1 + \exp(w_1 - w_2)\}^{-1}] = J_1(w).$$

The function $J(w)$ has bounded derivatives and therefore obeys the required Lipschitz condition. According to EM's Theorem 4, the model (2) is stable in the sense that ΔX has a stationary solution, if the spectral radius of the matrix

$$\begin{pmatrix} \Gamma & J_1(w) \\ \beta'\Gamma & I_2 + \beta' J_1(w) \end{pmatrix} \tag{3}$$

is less than $1 - \epsilon$.

For the model (2), estimation constitutes no problems, as the error-correction vector is given and thus OLS can be applied. Linear regression yields estimates $\hat{\mu}^*, \hat{\alpha}, \hat{\Gamma}$ of parameters μ^*, α, Γ, where μ^* denotes the total intercept $\mu - \alpha\delta$. Then $\hat{\delta}$ is obtained as the sample mean of the error-correction variable $\ln\{\exp(z_1 - y) + \exp(z_2 - y)\}$. In a second step, $\hat{\mu}$ is obtained as $\hat{\mu}^* + \hat{\alpha}\hat{\delta}$.

Unfortunately, this approach may yield estimated model structures that do not fulfil important features of the observed data. First, unrestricted $\hat{\alpha}$ may contain elements that violate stability conditions. A simple remedy is to replace such elements by zero. For example, $\hat{\alpha}_3$ may be negative, thus driving away the output variable y from the equilibrium. A valid model is obtained by replacing $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)'$ by $\hat{\alpha}^* = (\hat{\alpha}_1, \hat{\alpha}_2, 0)'$. This is not the maximum-likelihood estimate under the restriction $\alpha_3 = 0$, which would require a GLS–type correction.

Another critical feature of the OLS estimates may be inhomogeneous growth in X, whenever $\mathrm{E}(\Delta X)$ is not scalar. This does not impair nonlinear integratedness and stationarity of ΔX or of the error-correction variable. However, inhomogeneity tends to drive the ratios $\exp(c - y)$ and $\exp(e - y)$ to marginal values in the sense that, for the slower-growing component, for example c, $\exp(c-y)$ approaches 0 and $\exp(e-y)$ approaches $\exp(\delta)$.

Growth homogeneity can be imposed in the following steps. If the mean error-correction variable is δ, one obtains

$$E(\Delta X) = \mu + \Gamma E(\Delta X)$$

and therefore

$$E(\Delta X) = (\mathbf{I} - \Gamma)^{-1}\mu,$$

where $\mathbf{I}$ denotes the identity matrix. For example, the estimate $\hat{\mu}$ will lead to an estimate $m_x = (\mathbf{I} - \hat{\Gamma})^{-1}\hat{\mu}$ of $E(\Delta X)$. The averaged version

$$\tilde{m}_x = 3^{-1}(\mathbf{1}'m_x)\mathbf{1},$$

with $\mathbf{1} = (1, \ldots, 1)'$, then yields a modified estimate of μ as

$$\tilde{\mu} = (\mathbf{I} - \Gamma)\,\tilde{m}_x. \qquad (4)$$

The estimate $\tilde{\mu}$ enforces homogeneous growth among components and gives more realistic trajectories. Again, an alternative solution would be to use restricted maximum-likelihood estimation. If the generating model is simulated using the model (2) with an intercept $\mu^* = \mu - \alpha\delta$, then an estimate for this intercept is formed according to $\tilde{\mu}^* = \tilde{\mu} - \hat{\alpha}\hat{\delta}$.

3. THE DATA

We use seasonally adjusted quarterly national accounts data for the United Kingdom over the time range 1965–2005. Gross domestic product (GDP) in constant prices is used directly for the output series $\exp(y)$. Construction investment $\exp(c)$ results from adding the positions for gross fixed capital formation (GFCF) for housing and for other construction, while equipment investment $\exp(e)$ is obtained by adding the remaining GFCF components at constant prices. Due to the recent transition of national accounting from the traditional Paasche indices to chained deflators, the sum $\exp(c) + \exp(e)$ does not correspond exactly to the real total GFCF position of national accounting. The discrepancy is minor and disappears toward the end of the sample. We feel that using this sum $\exp(c) + \exp(e)$ for total investment is the most preferable option, in order to preserve the internal consistency of the data series.

Parallel analyses were conducted on three different data sets: Austrian data for the time range 1988–2002, French data 1970–1998, and an older version of U.K. accounts 1965–2003. For brevity, we will only mention the results from these investigations in brief.

Figure 1 shows the evolution over time of ratios of total GFCF and of the two main investment components to GDP. The ratio of total GFCF over GDP has remained fairly stable over the whole time range, at around 16%. By contrast, the

https://doi.org/10.1017/S1365100507060385 Published online by Cambridge University Press

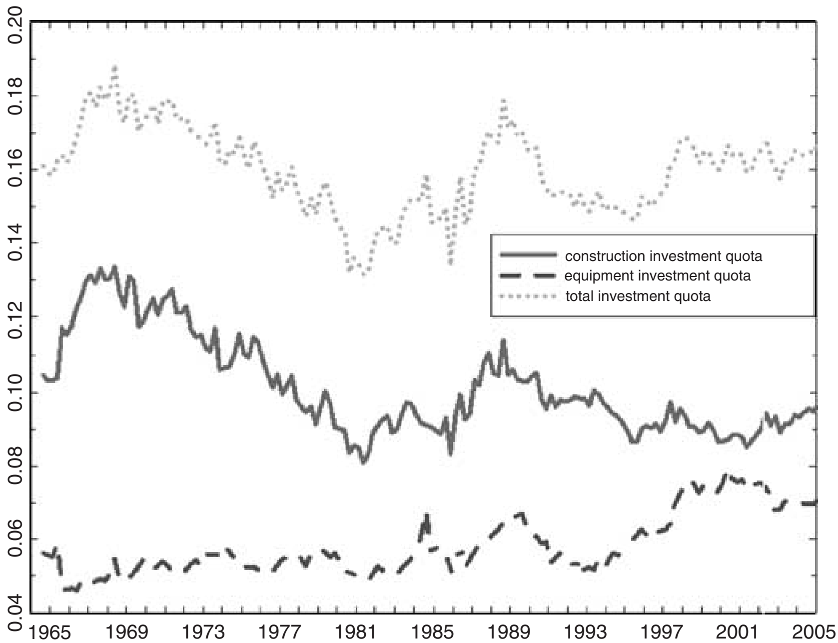

FIGURE 1. Shares of major investment components in British GDP. Quarterly data 1965–2005.

share of equipment investment has increased from around 5% to 7–8%, whereas the share of construction investment has fallen from a maximum of 13% in the 1960s to less than 10% over the same time range. Even more pronounced changes in their contribution appear when the shares of the original subaggregates of national accounts are analyzed. The construction of our two major components of investment follows the historical classification (Berndt 1996) that is still in predominant usage in many macroeconomic forecasting models.

In summary, the ratio of total investment to GDP is fairly stable over time, whereas construction investment shows a tendency to lose ground to the benefit of the rather heterogeneous aggregate of equipment (or, rather, "nonconstruction") investment. Similar tendencies were observed in the parallel data sets for Austria and France.

Some descriptive unit-root test statistics are summarized in Table 1. The variables are logarithms of the share of construction investment in GDP, of equipment investment in GDP, and of total GFCF in GDP. Whereas for the Dickey–Fuller tests the lag order was determined by the AIC information criterion, a window length of 4 was generally used for the Phillips–Perron version of the test. Both tests are unable to reject unit roots in the component ratios, whereas the unit root in the share of total GFCF is rejected at the 10% significance level. This result is in line with visual evidence.

TABLE 1. Unit root tests

	$c - y$	$e - y$	$g - y$
Dickey–Fuller tests			
Augmenting lags	4	7	5
μ-statistics	−1.669	−1.674	−2.672*
Phillips–Perron tests			
Window length	4	4	4
Statistics	−1.674	−1.646	−2.815*
RUR tests			
RUR	1.015**	1.562	1.640
RUR-BF	2.374	2.429	2.650

Note: c, e, g, and y are the logarithms of construction investment, equipment investment, total fixed investment, and gross domestic product, respectively. For the RUR test, simulated significance points are 1.11 and 3.56 at the 5% level, and 1.26 and 3.24 at the 10% level. For RUR-BF, simulated significance points are 1.73 and 4.92 at the 5% level, and 1.84 and 4.47 at the 10% level.

Additionally, we document the results of the recently introduced RUR (range unit root) test (Aparicio et al. 2006), which is designed to be more robust against outliers and nonlinearities. In contrast to the Dickey–Fuller and Phillips–Perron tests, it rejects its null of a unit root in its left tail if the variable is $I(0)$ without a trend, and in its right tail if the variable is $I(0)$ with a linear trend. By construction, the test may be sensitive to the presence of sublinear trends. Whereas the RUR test in its reportedly more powerful BF (backward and forward) version is unable to reject a unit root in all cases, its standard version rejects the unit root at the 5% level for the equipment investment ratio only. More experience with these new tests will be required before the information provided by this outcome can be properly assessed. However, the reported statistics sound a warning that the statistical evidence in favor of the theoretical concept—a stationary "great ratio" of total investment to GDP and time-changing subcomponent ratios—is far from unanimous. The results from several unreported bivariate and multivariate cointegration tests according to Johansen (1988, 1995) are also in line with the RUR tests. Unit-root tests on the parallel data sets yield ambiguous outcomes. Particularly for shorter time spans, the statistical tests were unable to reject a unit root for the total investment quota.

In summary, statistical support for the stationarity of the investment quota is unclear, although macroeconomic forecasters may tend to impose it for a longer-run prediction, for reasons of plausibility. By contrast, stationarity of subaggregate quotas is unsupported by nearly all statistics as well as by plausibility.

4. ESTIMATION

Table 2 gives the results of the preliminary unrestricted VAR estimation. Only one lag of differences was included according to information criteria. The nonlinear cointegration term has the correct sign for the investment subaggregates. For the

TABLE 2. Coefficient estimates: Estimation time range is 1965–2005.

	Δc_t	Δe_t	Δy_t
Δc_{t-1}	−0.176	−0.031	−0.013
	[−2.198]	[−0.301]	[−0.650]
Δe_{t-1}	0.073	−0.242	0.032
	[1.152]	[−3.002]	[1.974]
Δy_{t-1}	0.645	0.444	−0.066
	[1.935]	[1.047]	[−0.786]
μ^*	−0.146	−0.071	0.006
	[−1.912]	[−0.735]	[0.326]
α	−0.081	−0.042	0.000
	[−1.942]	[−0.805]	[0.008]
R^2	0.082	0.064	0.028

Note: Columns give the results for the three equations of the unrestricted nonlinear cointegration model.

output equation, it is very small and obtains the wrong sign for shortened samples that are used for forecasting, where it will be set to zero. For this restricted VAR system with nonlinear cointegration, mean forecasts with zero residuals are shown in Figure 2, whereas Figures 3 and 4 show stochastic forecasts. The low values of R^2 in Table 2 imply that the model has comparatively little explanatory power; hence the confidence bands are relatively wide.

The forecast for the total investment quota in Figure 3 reflects the fact that the British investment quota is only slightly above its historical mean at the end of the sample. Following a five-year decrease to that historical mean, the median forecast is constant over the remainder of the forecast time range. Figure 4 shows that construction investment is in a long-run decline that is assumed to continue into the future, whereas equipment investment is predicted to continue its relative increase. The distribution of the equipment predictors is asymmetric, and its upper 5% fractile even exceeds that of the investment predictor after 2022. The equipment median will overtake the construction median after 2030.

It is interesting to contrast this picture with the results from the older U.K. data set that ended in 2002. That data was much more supportive of an imminent overtaking of construction investment by equipment investment. Apparently, the long-run decline of the construction share has been weakening over the past few years.

If growth homogeneity is imposed on the British system (see Figure 5), median forecasts for both investment components become flat at approximate end-of-sample values of 7% for equipment and 9% for construction, thus reflecting the low degree of dynamic dependence in the model. This variant avoids the extremely high or low shares of the two subaggregates that evolve from the model with inhomogeneous growth, at the price of reducing the scenario to uninformative

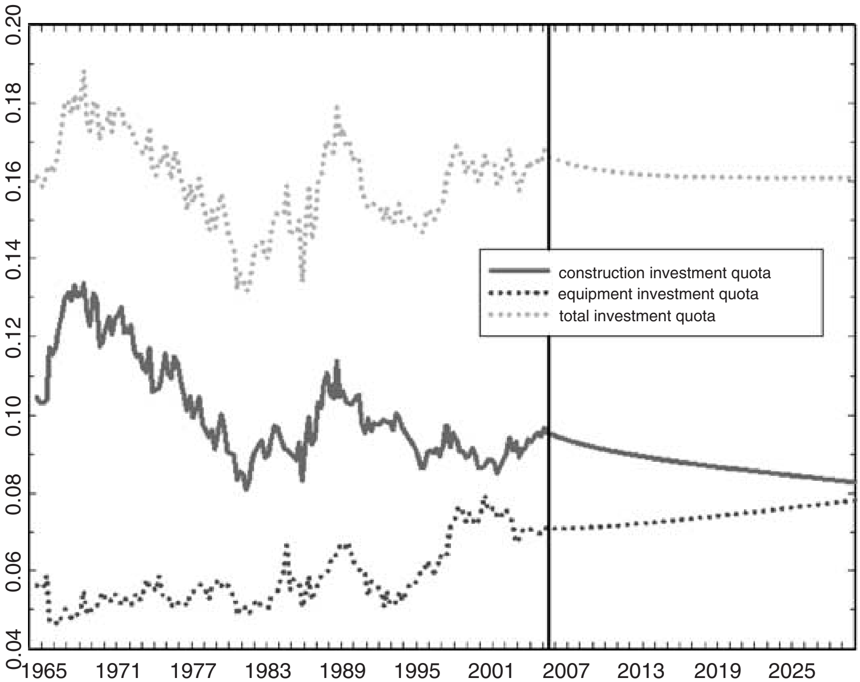

FIGURE 2. Twenty-five years of forecasting for the U.K. nonlinear system with zero residuals. A vertical bar separates the sample from the prediction interval.

random-walk behavior. Contrary to a usual cointegration model, there is no *stochastic* restriction on the difference between investment components, excepting the error-correction term for the sum, which keeps the components from undue expansion.

5. FORECASTING EVALUATION

5.1. Different Concepts of Evaluating Forecasts

Our intention is to rely mainly on visual evaluation of the plausibility of longer-range prediction rather than on a numerical prediction evaluation, because of the prominence of sample-specific features in any selected finite time range. However, for a better clarification of the issues, we also report the results of an out-of-sample prediction evaluation.

Regarding the stochastic assumption about the prediction model and the true model, one can distinguish four types of forecast evaluations. The simplest way of evaluating forecasts is by comparing a mean forecast x^*_{s+h} for an observation x_{s+h}, which is a function of the observations $x_1, \ldots, x_s$ and the observation x_{s+h}, typically by a distance function $g\left(x^*_{s+h}, x_{s+h}\right)$. Often, such evaluations are summarized by averaging over a range of values for s, tacitly assuming that

https://doi.org/10.1017/S1365100507060385 Published online by Cambridge University Press

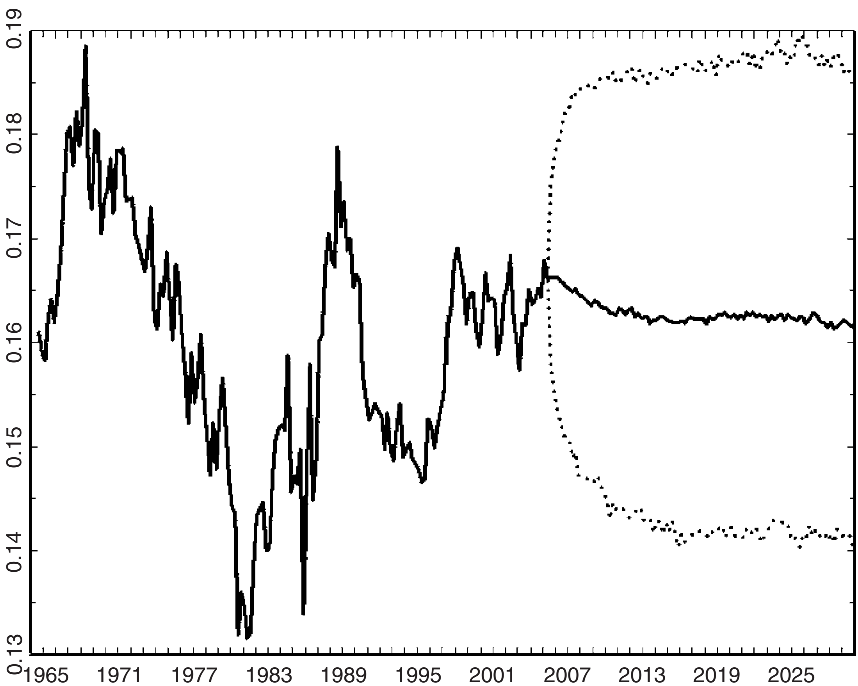

FIGURE 3. Median and upper and lower 5% quantiles for stochastic forecasts from the U.K. nonlinear system. Predicted variable is the total investment quota.

$m^{-1} \sum_{s=n-m+1}^{n} g(x^*_{s+h}, x_{s+h})$ converges to some constant, which we denote symbolically by $g(x^*, x)$. The forecasting model with the lowest value for $g(x^*, x)$ is then interpreted as the best one. This approach is not appropriate for nonlinear forecasting models or for stochastic prediction.

Most prediction experiments rely on a variant of approximating the integral

$$\int g(x^*, x) f(x)\, dx,$$

where x denotes the data-generating process, x^* is the forecast, and $g(x, y)$ is a distance function, for example, the squared distance $g_2(x, y) = (x - y)^2$. It is obvious that this approach is used as a backdrop for the prediction accuracy tests, as suggested by Diebold and Mariano (1995), for example. Again, for a stochastic forecast x^*, this interpretation of measuring accuracy is inconvenient, and the assumption of a true probability model for the data also appears to be restrictive.

As an alternative, one may consider integrals of the type

$$\int g(x^*, x) f(x^*)\, dx^*,$$

https://doi.org/10.1017/S1365100507060385 Published online by Cambridge University Press

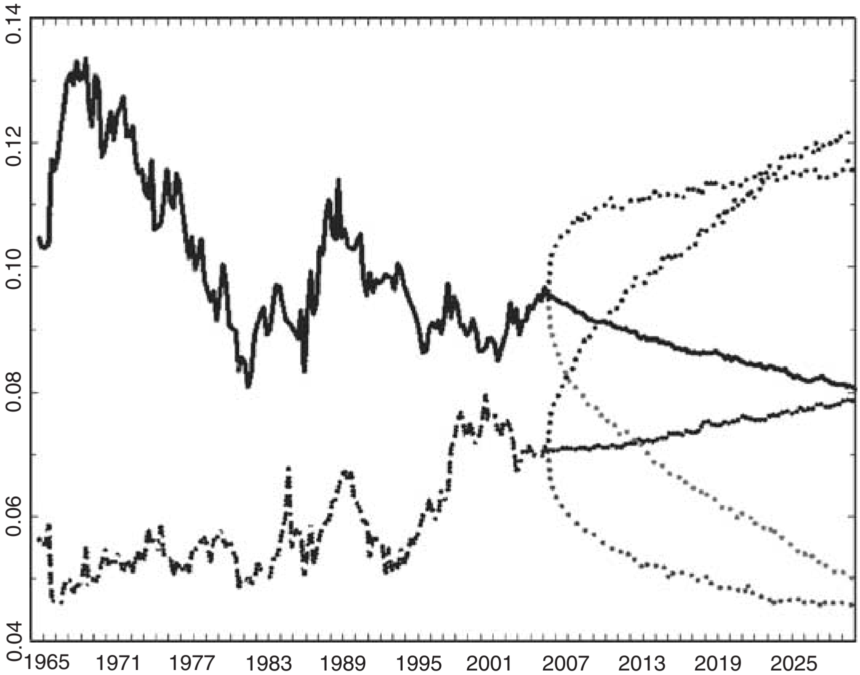

FIGURE 4. Stochastic prediction from the nonlinear model for the U.K. data. Median forecasts for construction investment quota (solid curve) and for equipment investment quota (dashed curve). Dotted curves represent lower and upper 5% quantiles of the forecast distribution.

where the expectation of $g(x^*, x)$ is conditioned on the observed data x and the forecast x^* is random. Such an interpretation is in line with the current popularity of fan charts and appears to be more appropriate for our purposes. For empirical applications, the integral is to be approximated by

$$\sum_{t=1}^{m} \sum_{\omega=1}^{\Omega} g(x_t^*(\omega), x_t),$$

where ω is drawn according to the probability distribution of the stochastic forecast. For our experiment, we use $\Omega = 200$, that is, there are 200 replications of the stochastic forecast, and $m = 50$, that is, we evaluate predictive accuracy over the last 12.5 years of the sample. We repeat the experiment for a whole range of horizons, ranging from single-step to 40-step forecasts, and remember to focus on a prediction horizon of 5 to 10 years. Note that for a horizon of h and a time range of m, only the first $n - h - m + 1$ observations can be used for estimating the model parameters, if the prediction is supposed to be truly out-of-sample.

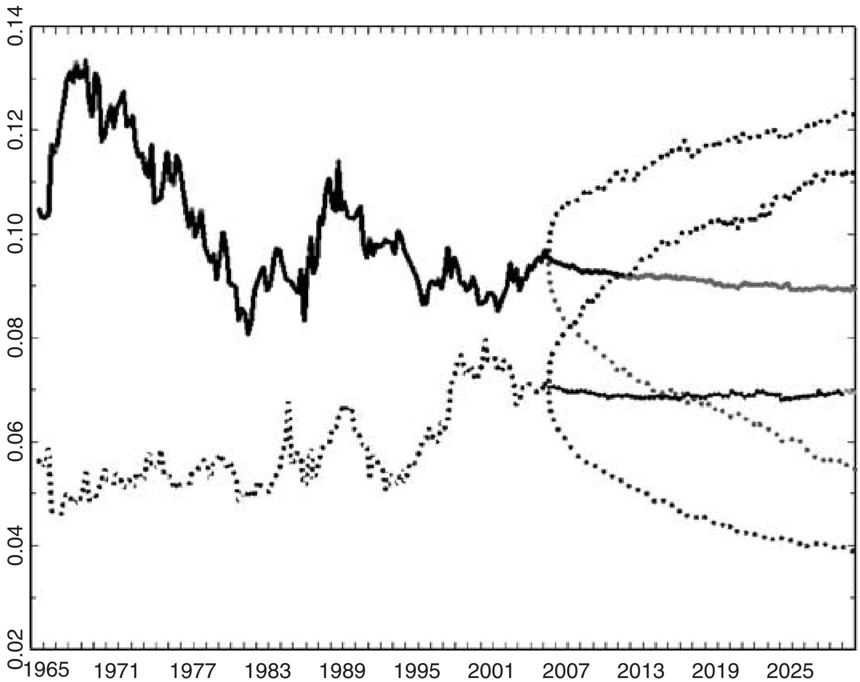

FIGURE 5. Stochastic prediction from the nonlinear model for the U.K. data under the restriction of growth homogeneity. Median forecasts for construction investment quota (solid curve) and for equipment investment quota (dashed curve). Dotted curves represent lower and upper 5% quantiles of the forecast distribution.

A fourth variant assumes stochastic processes for both the forecast x^* and the true model x, that is,

$$\int \int g(x^*, x) f(x^*, x)\, dx^*\, dx.$$

Here, the difficulty is that the generating law for x is unknown. A workable solution would be to act as if the estimated parameters from the observations $x_1, \ldots, x_s$ determined the true structure and to draw from the assumed statistical distribution. One obtains an approximation by sums of the form

$$\sum_{t=1}^{m} \sum_{\omega_1=1}^{\Omega_1} \sum_{\omega_2=1}^{\Omega_2} g[x_t^*(\omega_1), x_t(\omega_2)],$$

where the notation simplifies the fact that both the generating laws for x^* and for x are time-changing in the sense that an increasing training sample is used to determine the parametric structures. Such evaluations answer the question of whether the suggested prediction *method* performs satisfactorily, *if the true model*

https://doi.org/10.1017/S1365100507060385 Published online by Cambridge University Press

class is given. For example, the assumed true model class may be nonlinear cointegration models and the methods may be linear or nonlinear cointegration models. Due to sampling variation in parameter estimation, a match between method and true class does not necessarily define the best method. We shall first consider the above mentioned alternative with fixed x and then return to the double-stochastic version.

5.2. Evaluations Conditional on Observed Data

As candidates for stochastic forecasts, we use model-based predictions for five models: an unrestricted VAR in differences, the nonlinear cointegration model without and with imposing growth homogeneity, a linear VECM with stationary subcomponent ratios, and a VAR in differences with growth homogeneity. For all error-correction models, instability was excluded by replacing unstable influences of the error-correction terms to zero. This criterion was used separately at each time point, so that the experiment is out-of-sample in all regards. Although various other models could be used for a comparison, note that it is not necessary to impose growth homogeneity on the linear cointegration model, as it is fulfilled automatically.

For the function $g_2(x, y) = (x - y)^2$, that is, mean squared errors, results are displayed in Figures 6 and 7. For $g_1(x, y) = |x - y|$, that is, mean absolute errors, the ranking of forecasts is very similar. The benchmark models in differences without any further restriction clearly yield the largest prediction errors, and the effect of the growth homogeneity restriction is small. Contrary to the simulations of Engle and Yoo (1987), cointegrating models dominate at almost all horizons for all series, not only at larger horizons. Note, however, that we do not use the VAR in levels as a benchmark that was used by Engle and Yoo but in differences, and that we evaluate predictive accuracy for the (stationary or at least bounded) ratios and not for the assumedly integrated variables, such as y.

For the total investment quota and for the construction subaggregate, the nonlinear cointegration models dominate convincingly for longer prediction horizons. The linear cointegration model does well for short horizons but its performance deteriorates at the long end. For forecasting the equipment investment quota, the assumption of linear cointegration of subcomponents and therefore also of the equipment quota yields the optimum results. The performance of all rival models is similar, with a slight lead of the nonlinear model with growth homogeneity. Presumably, growth homogeneity would prove more beneficial at even longer horizons, whereas different growth rates for subaggregates are acceptable within the limits of the experiment. The general impression from all three variables is that the nonlinear cointegration model with growth homogeneity yields the most satisfactory results.

Similar results were obtained for the French and Austrian data sets, as well as for an older variant of the British data. For most series, always including the total investment quota, the nonlinear cointegration model achieves the best accuracy,

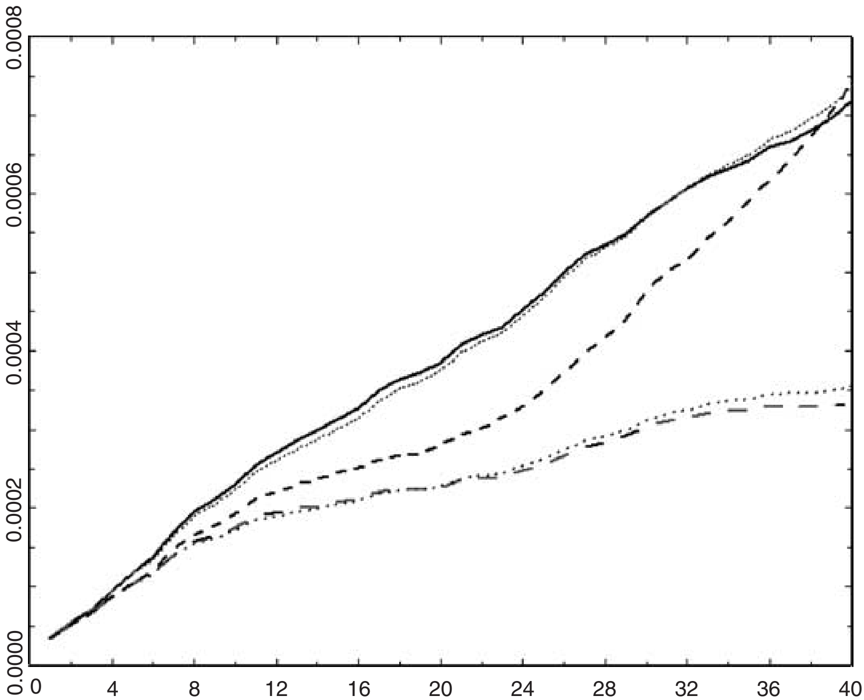

FIGURE 6. Mean squared errors for prediction horizons $h = 1$ to $h = 40$. Predicted variable is the U.K. construction investment quota. Curves represent VAR in differences (solid), nonlinear cointegration model (long dashes), nonlinear cointegration model with growth homogeneity (dots), linear cointegration of subcomponents (short dashes), and VAR in differences with growth homogeneity (fine dots).

and it typically benefits from imposing growth homogeneity at longer horizons. While the linear cointegration model is competitive for some component series, it tends to be less attractive for the total. The models in differences rank last, at least for longer horizons.

5.3. Evaluations Conditional on Simulated Data

These evaluations assume that a specified model class is the correct one and determines the free parameters by estimation from the full available sample. From this estimated pseudo-true model, artificial samples are generated ("parametric bootstrap"), which are then "predicted" using all of the previously specified methods. One of the methods corresponds to the class used for generating the data. These evaluations are helpful, as they provide information on the relative merits with regard to the accuracy of forecasts from correctly specifying the model class. Because of sampling variation in parameter estimation, the true model class is not necessarily the best one at all forecast horizons.

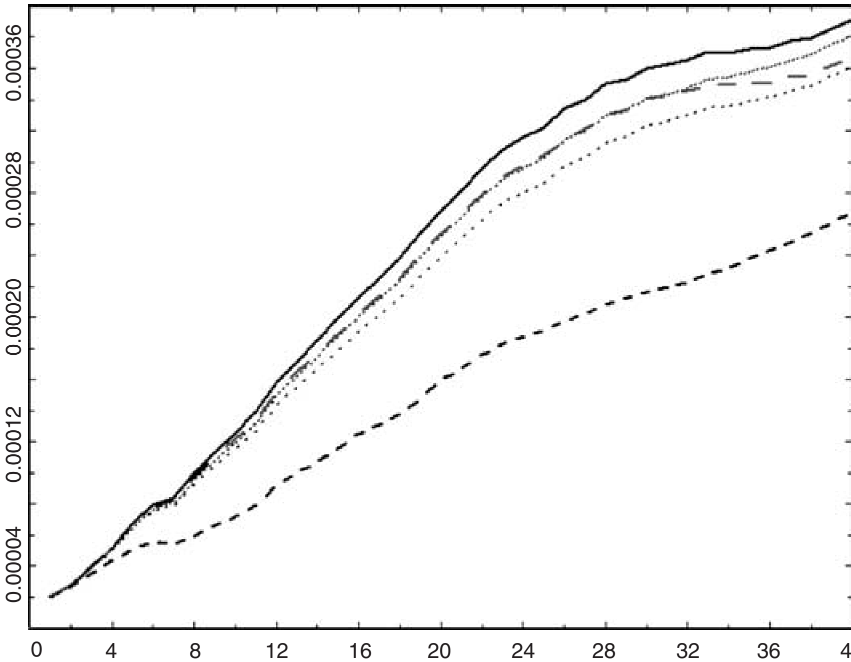

FIGURE 7. Mean squared errors for prediction horizons $h = 1$ to $h = 40$. Predicted variable is the U.K. equipment investment quota. Curves represent VAR in differences (solid), nonlinear cointegration model (long dashes), nonlinear cointegration model with growth homogeneity (dots), linear cointegration of subcomponents (short dashes), and VAR in differences with growth homogeneity (fine dots).

Figures 8 and 9 rely on experiments using 100 replications both for the stochastic predictors and for the pseudo-true model. The assumed true model class is the nonlinear cointegration model with the growth homogeneity restriction. Predictions based on the true model dominate at all horizons, while the ranking of the other predictors varies. For the equipment investment quota, the nonlinear model without the homogeneity restriction falls behind the linear model in pure differences with the homogeneity restriction, whereas for the construction and total quotas, the unrestricted nonlinear model comes in at a clear second position. The comparatively good performance of the linear cointegration models in the observed data is not matched in the simulated data.

These simulations offer an informal test of whether the assumed model is a likely data-generating mechanism for the observed data, even though such tests are not in the focus of our investigation. If a nonlinear cointegration model actually had generated the British investment data, Figures 8 and 9 should roughly match the features seen in the observational counterparts, Figures 6 and 7. Although that correspondence is acceptable in general, there are some noteworthy differences. First, although the empirical plots support the linear cointegration model

https://doi.org/10.1017/S1365100507060385 Published online by Cambridge University Press

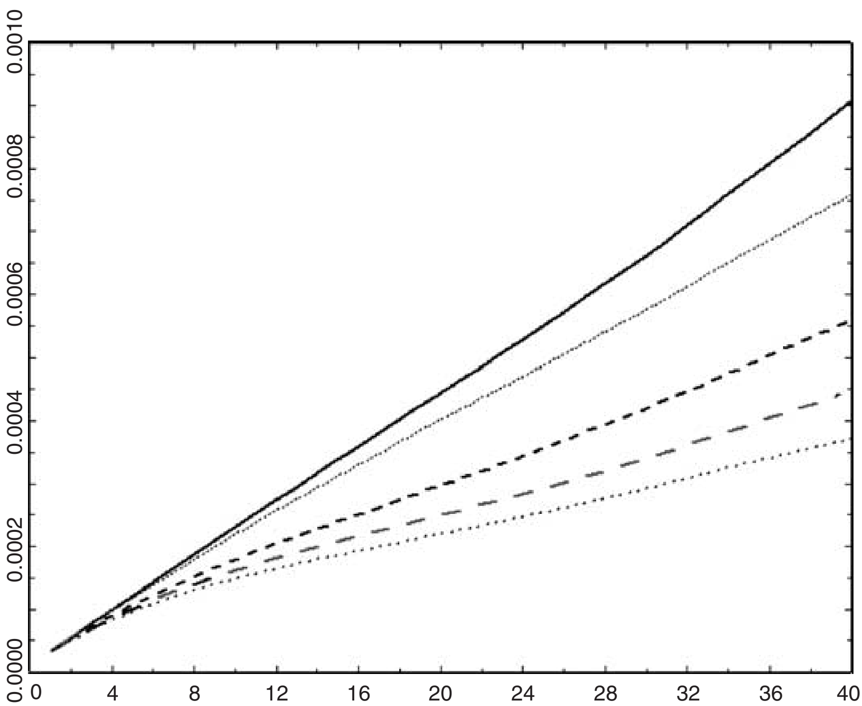

FIGURE 8. Mean squared errors for prediction horizons $h = 1$ to $h = 40$. Predicted variable is a parametric bootstrap version of the British construction investment quota. Curves represent VAR in differences (solid), nonlinear cointegration model (long dashes), nonlinear cointegration model with growth homogeneity (dots), linear cointegration of subcomponents (short dashes), and VAR in differences with growth homogeneity (fine dots).

as a forecasting tool at some prediction horizons, this model is not among the preferred ones for the simulation graphs. This mismatch may indicate that true data behavior is in between the linear and the nonlinear model, in the sense that the persistence of subcomponent quotas is stronger than would be implied by the nonlinear cointegration model, though not as strong as would be implied by the linear model. Second, the ranking of prediction models is robust to the prediction horizon for the bootstrap version, whereas it critically depends on the horizon for the empirical version. This may indicate that longer-run cycles play a larger role in empirical data than in all suggested model classes. These longer-run cycles may reflect cycles in political attitudes, as particularly construction investment is severely influenced by policy decisions. Finally, the numerical values of the mean squared errors show noteworthy differences, which however is to be expected due to sampling variation, if the data is viewed as a single observation of a trajectory from a time-series process.

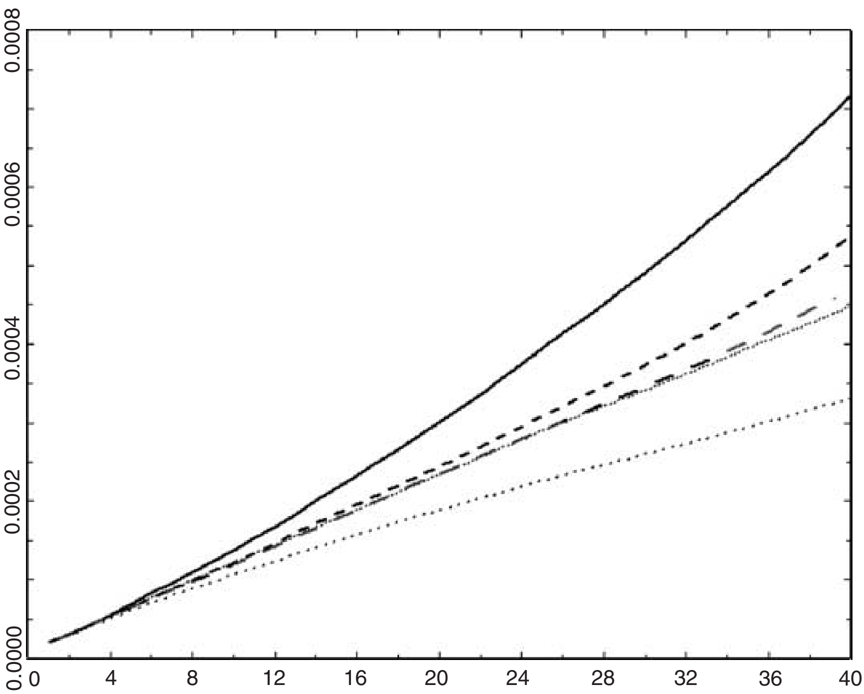

FIGURE 9. Mean squared errors for prediction horizons $h = 1$ to $h = 40$. Predicted variable is a parametric bootstrap version of the British equipment investment quota. Curves represent VAR in differences (solid), nonlinear cointegration model (long dashes), nonlinear cointegration model with growth homogeneity (dots), linear cointegration of subcomponents (short dashes), and VAR in differences with growth homogeneity (fine dots).

The parametric bootstrap evaluation can be repeated for any other assumed data-generating process. For brevity, we summarize the results of such experiments without reporting details. Generally, we find that the data-generating model class does not always dominate with regard to predictive accuracy.

For example, imposing growth homogeneity tends to improve forecasts, even when the true model does not obey the homogeneity condition. This observation is in line with the traditional recommendation of using simplified structures in forecasting, as parameter estimates with large uncertainty tend to deteriorate performance, even in cases where the parameters are statistically significant. Although this effect necessarily disappears as the sample size increases, it is definitely present for sample sizes of empirical interest in macroeconomics.

Similarly, simulating a linear cointegration structure that has been fitted to the actual data entails that the nonlinear model with growth homogeneity dominates as a forecasting tool. Without homogeneity being imposed, the model does not perform as well, but even then it beats the true data-generating model class for some horizons and series. We note that the linear cointegration model has more parameters to be estimated, and some of the adjustment coefficients are relatively imprecise. Ironically, the seemingly complex nonlinear model is simpler than the linear rival, and therefore outperforms it.

Similar results were obtained for simulations that built on the Austrian and French data. The nonlinear model with growth homogeneity yields the most robust forecasting structure, whereas the models in pure differences without imposing any long-run equilibrium conditions fail.

Finally, the technique of parametric bootstrap permits a review of the unit-root statistics reported in Table 1. If the nonlinear model with homogeneity restriction indeed generated the data, we would expect test statistics in the range of the reported values, whether those point literally to unit roots or not. This is indeed the case. The bootstrap processes yield Dickey–Fuller statistics with a median of -0.66 to -1.18 for the subcomponent ratios and of -2.36 for the total ratio. The sample values are slightly to the left of these values for the subcomponent ratios but still not in the tails in the distribution, whereas the sample value for the total ratio achieves a close match.

Similarly, the medians of the distributions of the bootstrapped RUR test statistics are in the range of 1.33 to 1.72 and of 1.82 to 2.37 for the RUR-BF variant. These medians roughly match the sample values in Table 2, which implies that the RUR tests are unable to reject the $I(1)$ hypothesis for variables that have been generated by the nonlinear model and are otherwise similar to the observed data.

In summary, the apparent mismatch between the statistical properties of the suggested nonlinear process and the test results reported in Table 1 can be explained fully by the particular features of the nonlinear model and by some additional data-specific short-run parameters and starting values. Safe statistical detection of nonlinearities of the suggested type would require either a hitherto unknown test statistic or an implausibly long sample. Our bootstrap technique provides some insight into this problem and does not exclude the possibility that the actual data-generating process may be close to the suggested nonlinear structure.

6. SUMMARY AND CONCLUSION

We argue that, although macroeconomic theory postulates constancy of the share of total investment in output in the steady state of an economy, there is little information on the development of its components. We build on this argument by means of a trivariate time-series model for forecasting.

The task of forecasting the subcomponents, under the restriction of long-run constancy of the total investment share, yields a relatively simple example for nonlinear cointegration. We suggest the elimination of unstable features from longer-run scenarios in an iterative dialog with the data. Particularly in longer-run forecasting, statistical test decisions obtained from samples of limited length should be overridden in favor of admissibility restrictions imposed by plausibility and economic theory. Our results show that automatically generated forecasts from statistical analysis of linear structures may not fulfil such plausibility requirements. In scenarios with a very long time horizon, restricting the deterministic part of the system can also become crucial in order to avoid that one subcomponent disappears in the longer run.

We applied our nonlinear cointegration modelling idea to a three-variable system for two main investment components and aggregate output in the United Kingdom. Parallel experiments were performed on Austrian and French data.

Our main finding is that the nonlinear cointegration model is superior as a forecasting device at the projection horizons of interest of up to 10 years. We confirm the evidence through parametric bootstrapping that assumes the nonlinear model as the true model. Interestingly, the nonlinear cointegration model even dominates when the experiment assumes that the DGP conforms to a linear cointegration model.

The presented partial models on investment and output can be used as building blocks in larger macroeconometric models, where the relationships between the investment sector and other sectors of the economy can be fully captured. In order to permit a focus on the main issues, we exclude such extensions from the present paper. The study could, however, form a basis for future work on this subject.

Without substantial modification, techniques similar to the one outlined in this paper can be used for other cases where the total share is known to be more constant than the shares of components, such as in modeling components of consumer demand. We feel, however, that the basic modeling idea may have even wider applicability in other areas of economic modeling.

REFERENCES

Aparicio, Felipe, Alvaro Escribano, and Ana E. Sipols (2006) Range unit-root (RUR) tests: Robust against nonlinearities, error distributions, structural breaks and outliers. *Journal of Time Series Analysis* 27, 545–576.

Berndt, Ernst R. (1996) *The Practice of Econometrics: Classic and Contemporary.* Boston: Addison-Wesley.

Christoffersen, Peter F. and Francis X. Diebold (1998) Cointegration and long-horizon forecasting. *Journal of Business & Economics Statistics* 16, 450–458.

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

Engle, Robert F. and Byung S. Yoo (1987) Forecasting and testing in co-integrated systems. *Journal of Econometrics* 35, 143–159.

Escribano, Alvaro and Santiago Mira (2002) Nonlinear error correction models. *Journal of Time Series Analysis* 23, 509–522.

Escribano, Alvaro (2004) Nonlinear error correction: The case of money demand in the UK (1878–2000). *Macroeconomic Dynamics* 8, 76–116.

Johansen, Soren (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Johansen, Søren (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson (1991) Stochastic trends and economic fluctuations. *American Economic Review* 81, 819–840.
Klein, Lawrence R. and Richard F. Kosobud (1961) Some econometrics of growth: Great ratios of economics. *Quarterly Journal of Economics* 75, 173–198.
Kunst, Robert M. and Klaus Neusser (1990) Cointegration in a macroeconomic system. *Journal of Applied Econometrics* 5, 351–365.
Romer, David (1996) *Advanced Macroeconomics*. New York: McGraw-Hill.
Stock, James H. and Mark W. Watson (1988) Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.