

7

Seeing into the Future

Abstract While empirical modelling is primarily concerned with understanding the interactions between variables to recover the underlying ‘truth’, the aim of forecasts is to generate useful predictions about the future regardless of the model. We explain why models must be different in non-stationary processes from those that are optimal’ under stationarity, and develop forecasting devices that avoid systematic failure after location shifts.

Keywords Forecasting · Forecast failure · Forecast uncertainty · Hedgehog forecasts · Outliers · Location shifts · Differencing · Robust devices

In a stationary world, many famous theorems about how to forecast optimally can be rigorously proved (summarised in Clements and Hendry 1998):

1. causal models will outperform non-causal (i.e., models without any relevant variables);
2. the conditional expectation of the future value delivers the minimum mean-square forecast error (**MSFE**);

3. mis-specified models have higher forecast-error variances than correctly specified ones;
4. long-run interval forecasts are bounded above by the unconditional variance of the process;
5. neither parameter estimation uncertainty nor high correlations between variables greatly increase forecast-error variances.

Unfortunately, when the process to be forecast suffers from location shifts and stochastic trends, and the forecasting model is mis-specified, then:

1. non-causal models can outperform correct in-sample causal relationships;
2. conditional expectations of future values can be badly biased if later outcomes are drawn from different distributions (see Fig. 4.5);
3. the correct in-sample model need not outperform in forecasting, and can be worse than the average of several devices;
4. long-run interval forecasts are unbounded;
5. parameter estimation uncertainty can substantively increase interval forecasts; as can
6. changes in correlations between variables at or near the forecast origin.

The problem for empirical econometrics is not a plethora of excellent forecasting models from which to choose, but to find any relationships that survive long enough to be useful: as we have emphasized, the stationarity assumption must be jettisoned for observable variables in economics. Location shifts and stochastic trend non-stationarities can have pernicious impacts on forecast accuracy and its measurement: Castle et al. (2019) provide a general introduction.

7.1 Forecasting Ignoring Outliers and Location Shifts

To illustrate the issues, we return to the two data sets in Chapter 5 which were perturbed by an outlier and a location shift respectively, then modelled by IIS and SIS. The next two figures use the indicators found in those examples. In Fig. 7.1, the 1-step forecasts with and without the indicator show the former to be slightly closer to the outcome, and with a smaller interval forecast.

Both features seem sensible: an outlier is a transient perturbation, and providing it is not too large, its impact on forecasts should also be transient and not too great. The increase in the interval forecast is due to the rise in the estimated residual standard error from the outlier. Nevertheless, failing to model outliers can be very detrimental as Hendry and Mizon (2011) show when modelling an extension of the US food expenditure data noted above, which was of course, the origin of IIS finding the very large outliers in the 1930s, discussed in Sect. 5.1 as a robust estimation method.

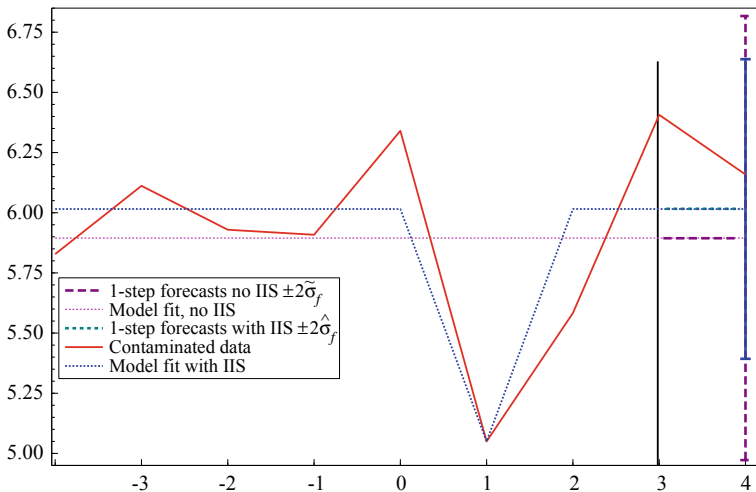


Fig. 7.1 1-step forecasts with and without the impulse indicator to model an outlier

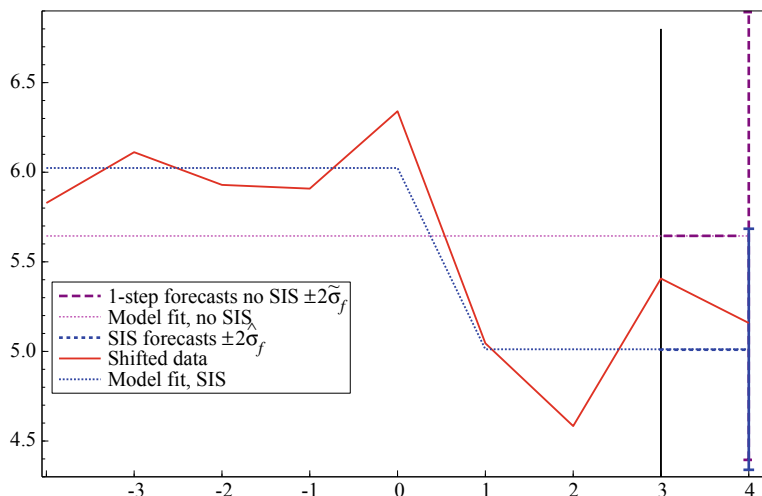


Fig. 7.2 1-step forecasts with and without the step indicator

However, the effect of omitting a step indicator that matches a location shift is far more serious as Fig. 7.2 shows. The 1-step forecast with the indicator is much closer to the outcome, with an even smaller interval forecast than that from the model without. Moreover, the forecast without the step indicator is close to the top of the interval forecast from the model with.

In Fig. 7.2, we (the writers of this book) know that the model with SIS matches the DGP (albeit with estimated rather than known parameters), whereas the model that ignores the location shift is mis-specified, and its interval forecast is hopelessly too wide—wider than the range of all previous observations. Castle et al. (2017) demonstrate the use of SIS in a forecasting context, where the step-indicator acts as a type of intercept correction when there has been a change in policy resulting in a location shift. An intercept correction changes the numerical value of the intercept in a forecasting model by adding a recent forecast error to put the forecast ‘back on track’. SIS, along with other forms of robust device such as a conventional intercept correction, can greatly improve forecasts when they are subject to shifts at or near the forecast origin: see Clements and Hendry (1996).

7.2 Impacts of Stochastic Trends on Forecast Uncertainty

Because $I(1)$ processes cumulate shocks, even using the correct in-sample model leads to much higher forecast uncertainty than would be anticipated on $I(0)$ data. This is exemplified in Fig. 7.3 showing multi-period forecasts of $\log(\text{GDP})$ starting in 1990 till 2030: the outcomes to 2016 are shown, but not used in the forecasts. Constant-change, or difference stationary, forecasts (dotted) and deterministic trend forecasts (dashed) usually make closely similar central forecasts as can be seen here. But deterministic linear trends do not cumulate shocks, so irrespective of the data properties, and hence even when the data are actually $I(1)$, their uncertainty is measured as if the data were stationary around the trend.

Although the data properties are the same for the two models in Fig. 7.3, their estimated forecast uncertainties differ dramatically (bars and bands respectively), increasingly so as the horizon grows, due to the linear trend model assuming stable changes over time. Thus, model

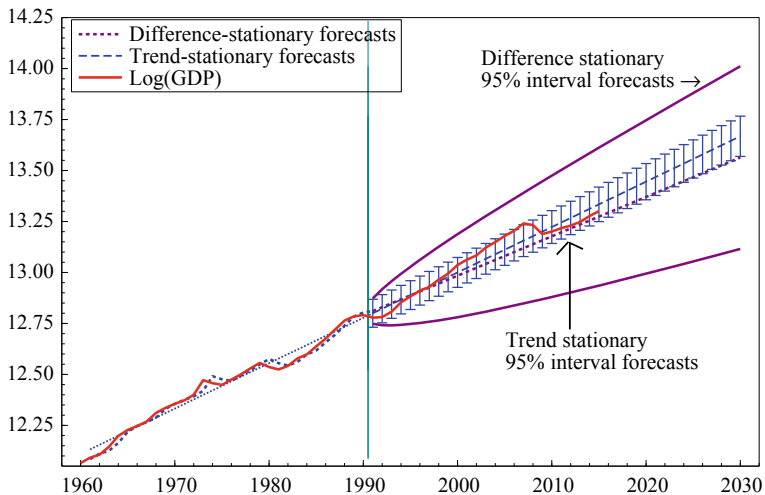


Fig. 7.3 Multi-period forecasts of $\log(\text{GDP})$ using a non-stationary stochastic-trend model (dotted) and a trend-stationary model (dashed) with their associated 95% interval forecasts

choice has key implications for measuring forecast uncertainty, where mis-specifications—such as incorrectly imposing linear trends—can lead to understating the actual uncertainty in forecasts. Although the assumption of a constant linear trend is rarely satisfactory, nevertheless, here almost all the outcomes between 1990 and 2016 lie within the bars. Conversely, the difference stationary interval forecasts are very wide. In fact, that model has considerable residual autocorrelation which the bands do not take into account, so over-estimate the uncertainty. However, caution is always advisable when forecasting integrated time series for long-periods into the future by either approach, especially from comparatively short samples.

7.3 Impacts of Location Shifts on Forecast Uncertainty

Almost irrespective of the forecasting device used, forecast failure would be rare in a stationary process, so episodes of forecast failure confirm that many time series are not stationary. Conversely, forecasting in the presence of location shifts can induce systematic forecast failure, unless the forecasting model accounts for the shifts.

Figure 7.4 shows some recent failures in 8-quarter ahead forecasts of US log real GDP. There are huge forecast errors (measured by the vertical distance between the forecast and the outcome), especially at the start of the ‘Great Recession’, which are not corrected till near the trough. We call these ‘hedgehog’ graphs since the successively over-optimistic forecasts lead to spikes like the spines of a hedgehog. It can be seen that the largest and most persistent forecast errors occur after the trend growth of GDP slows, or falls. This is symptomatic of a fundamental problem with many model formulations, which are equilibrium-correction mechanisms (EqCMs) discussed in Sect. 4.2: they are designed to converge back to the previous equilibrium or trajectory. Consequently, even when the equilibrium or trajectory shifts, EqCMs will persistently revert to the old equilibrium—as the forecasts in Fig. 7.4 reveal—until either the model is revised or the old equilibrium returns.

Figure 7.4 illustrates the difficulties facing forecasting deriving from wide-sense non-stationarity. However, the problem created by a location

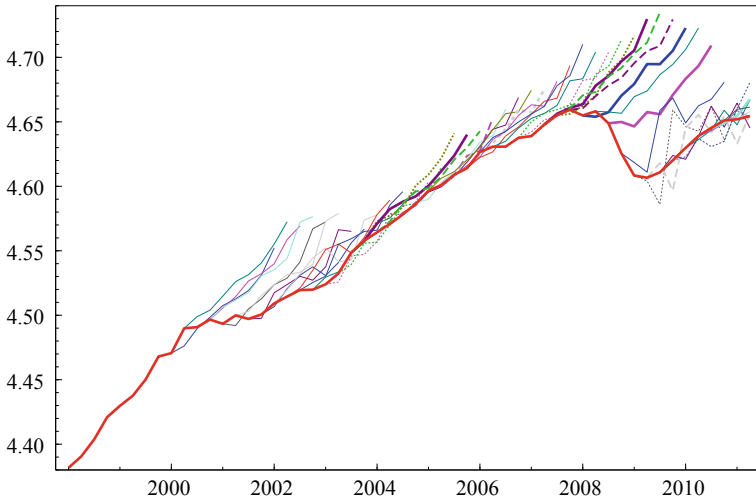


Fig. 7.4 US real GDP with many successive 8-quarter ahead forecasts

shift is not restricted to large forecast errors, but also affects the formation of expectations by economic actors: in theory models, today's expectation of tomorrow's outcome is often based on the 'most likely outcome', namely the conditional expectation of today's distribution of possible outcomes. In processes that are non-stationary from location shifts, previous expectations can be poor estimates of the next period's outcome. Figure 4.5 illustrated this problem, which has adverse implications for economic theories of expectations based on so-called 'rational' expectations. This issue also entails that many so-called structural econometric models constructed using mathematics based on inter-temporal maximization behavioural assumptions are bound to fail when the distributions involved shift as shown in Sect. 4.4.

7.4 Differencing Away Our Troubles

Differencing a break in a trend results in a location shift, as can be seen in Fig. 7.5, and in turn differencing a location shift produces an impulse, and a final differencing creates a 'blip'. All four types occur empirically.

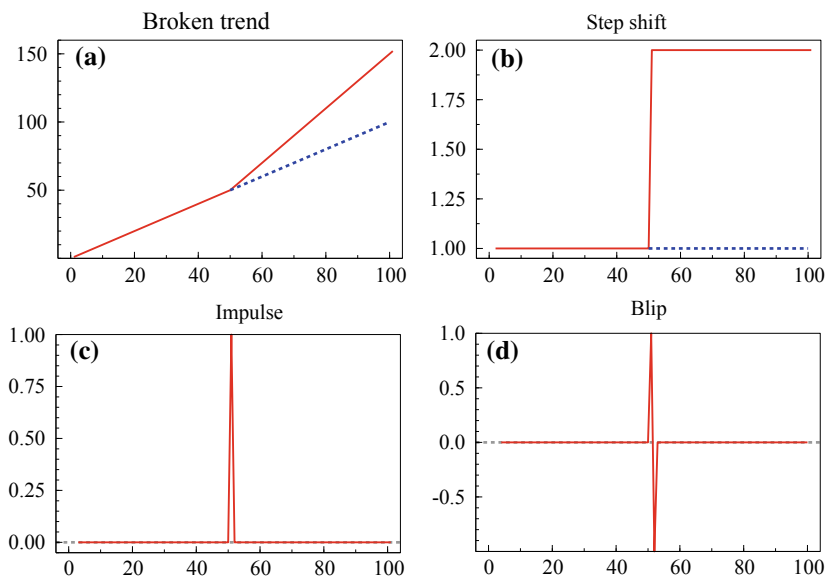


Fig. 7.5 Successively differencing a trend break in (a) creates a step shift in (b) an impulse in (c) and a 'blip' in (d)

Failing to allow for trend breaks or location shifts when forecasting entails extrapolating the wrong values and can lead to systematic forecast failure as shown by the dotted trajectories in Panels (a) and (b). However, failing to take account of an impulse or a blip just produces temporary errors, so forecasts revert back to an appropriate level rapidly. Consequently, many forecasts are reported for growth rates and often seem reasonably accurate: it is wise to cumulate such forecasts to see if the entailed levels are correctly predicted.

Figure 7.6 illustrates for artificial data: only a couple of the growth-rate outcomes lie above the 95% interval forecasts, but the levels forecasts are systematically downward biased from about observation 35. This is because the growth forecasts are on average slightly too low, which cumulates over time. The graphs show multi-step forecasts, but being simply a constant growth-rate forecast, the same interval forecasts apply at all steps ahead.

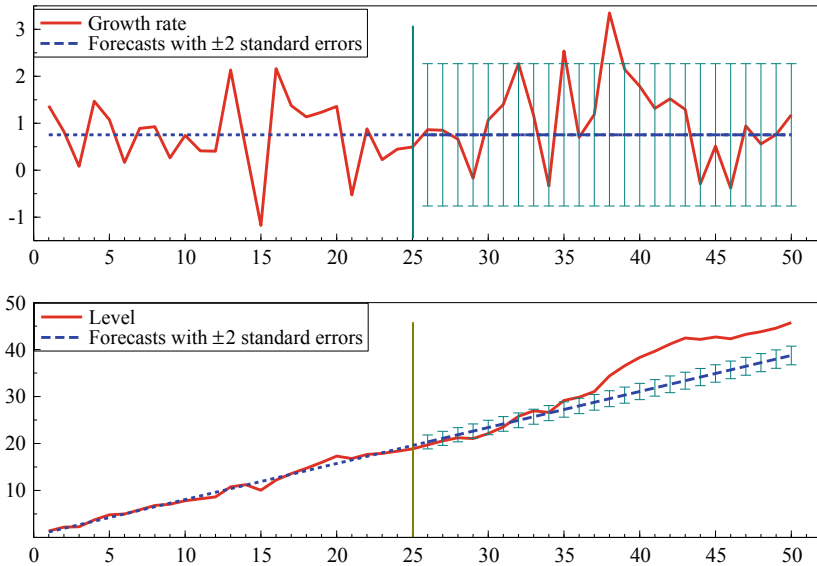


Fig. 7.6 Top panel: growth-rate forecasts; lower panel: implied forecasts of the levels

Constant growth-rate forecasts are of course excellent when growth rates stay at similar levels, but otherwise are too inflexible. An alternative is to forecast the next period's growth rate by the current value, which is highly flexible, but imposes a unit root even when the growth rate is $I(0)$. Figure 7.3 contrasted deterministic trend forecasts with those from a stochastic trend, which had huge interval forecasts. Such intervals correctly reflect the ever increasing uncertainty arising from cumulating unrelated shocks when there is indeed a unit root in the DGP.

However, forecasting an $I(0)$ process by a unit-root model also leads to calculating uncertainty estimates like those of a stochastic trend: the computer does not know the DGP, only the model it is fed. We must stress that interval forecasts are based on formulae that are calculated for the model used in forecasting. Most such formulae are derived under the assumption that the model is the DGP, so can be wildly wrong when that is not the case.

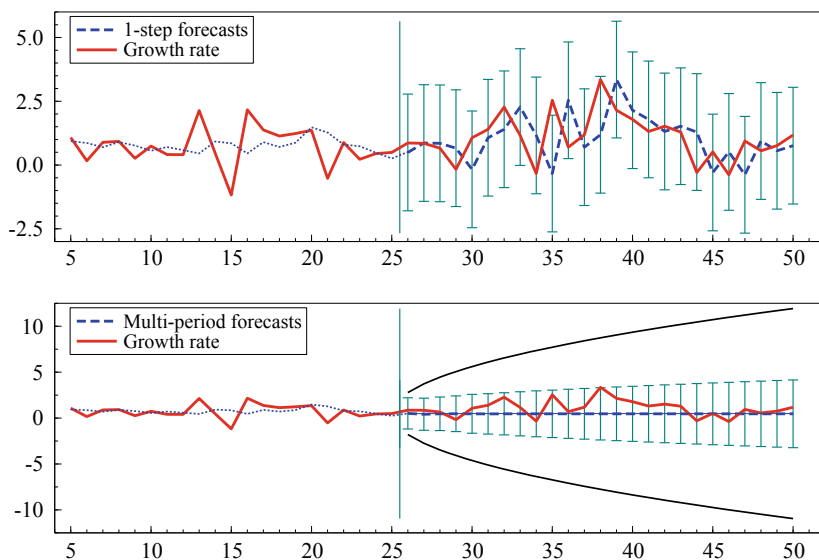


Fig. 7.7 Top panel: 1-step growth-rate forecasts from a 4-period moving average; lower panel: multi-period growth-rate forecasts with ± 2 standard errors from a random walk (bands) and a 4-period moving average of past growth rates (bars)

The top panel in Fig. 7.7 shows that 1-step growth-rate forecasts from a 4-period moving average of past growth rates with an imposed unit coefficient are much more flexible than the assumed constant growth rate, and only one outcome lies outside the 95% error bars. The two sets of multi-period interval forecasts in the lower panel of Fig. 7.7 respectively compare the growth rate and the 4-period moving average of past growth rates as their sole explanatory variables, both with an imposed unit coefficient to implement a stochastic trend. The average of the four most recent growth rates at the forecast origin, as against just one, produces a marked reduction in the interval forecasts despite still cumulating shocks.

A potential cost is that it will take longer to adjust to a shift in the growth rate. Here the growth rate is an $I(0)$ variable, and it is the imposition of the unit coefficient that creates the increasing interval forecasts, but even so, the averaging illustrates the effects of smoothing. This idea of smoothing applies to the robust forecasting methods noted in the next section. Care is required in reporting interval forecasts for several steps ahead as their

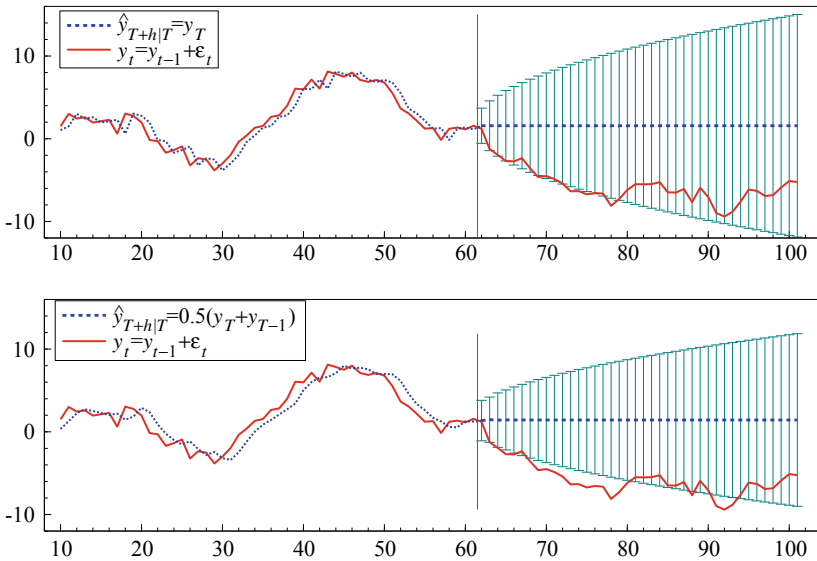


Fig. 7.8 Top panel: multi-period forecasts with ± 2 standard errors from the DGP of a random walk; lower panel: multi-period forecasts from a 2-period moving average with ± 2 calculated standard errors

calculation may reflect the properties of the model being used more than those of the DGP.

Conversely, trying to smooth a genuine random walk process by using a short moving average to forecast can lead to forecast failure as Fig. 7.8 illustrates. The DGP is the same in both panels, but the artificially smoothed forecasts in the lower panel have too small calculated interval forecasts.

7.5 Recommendations When Forecasting Facing Non-stationarity

Given the hazards of forecasting wide-sense non-stationary variables, what can be done? First, be wary of forecasting $I(1)$ processes over long time horizons. Modellers and policy makers must establish when they are dealing with integrated series, and acknowledge that forecasts then entail increas-

ing uncertainty. The danger is that uncertainty can be masked by using mis-specified models which can falsely reduce the reported uncertainty. An important case noted above is enforcing trend stationarity, as seen in Fig. 7.3, greatly reducing the measured uncertainty without reducing the actual, a recipe for poor policy and intermittent forecast failure. As Sir Alex Cairncross worried in the 1960s: 'A trend is a trend is a trend, but the question is, will it bend? Will it alter its course through some unforeseen force, and come to a premature end?' Alternatively, it is said that the trend is your friend till it doth bend.

Second, once forecast failure has been experienced, detection of location shifts (see Sect. 4.5) can be used to correct forecasts even with only a few observations, or alternatively it is possible to switch to more robust forecasting devices that adjust quickly to location shifts, removing much of any systematic forecast biases, but at the cost of wider interval forecasts (see e.g., Clements and Hendry 1999).

Nevertheless, we have also shown that one aspect of the explosion in interval forecasts from imposing an integrated model after a shift in an $I(0)$ process (i.e., one that does not have a genuine unit root) is due to using just the forecast-origin value, and that can be reduced by using moving averages of recent values. In turbulent times, such devices are an example of a method with no necessary verisimilitude that can outperform the in-sample previously correct representation. Figure 7.9 illustrates the substantial improvement in the 1-step ahead forecasts of the log of UK GDP over 2008–2012 using a robust forecasting device compared to a 'conventional' method. The robust device has a much smaller bias and MSFE, but as it is knowingly mis-specified, clearly does not justify selecting it as an economic model—especially not for policy.

That last result implies that it is important to refrain from linking out-of-sample forecast performance of models to their 'quality' or verisimilitude. When unpredictable location shifts occur, there is no necessary link between forecast performance and how close the underlying model is to the truth. Both good and poor models can forecast well or badly depending on unanticipated shifts.

Third, the huge class of equilibrium-correction models includes almost all regression models for time series, autoregressive equations, vector autoregressive systems, cointegrated systems, dynamic-stochastic general

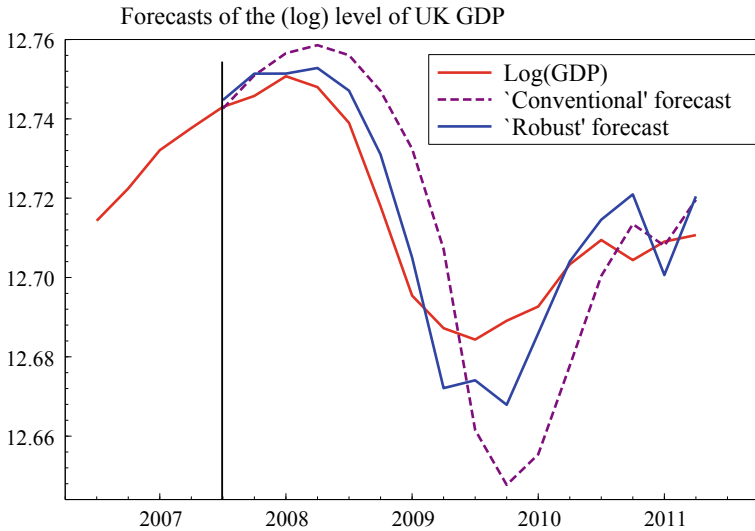


Fig. 7.9 1-step ahead forecasts of the log of UK GDP over 2008–2012 by 'conventional' and robust methods

equilibrium (DSGE) models, and many of the popular forms of model for autoregressive heteroskedasticity (see Engle 1982). Unfortunately, all of these formulations suffer from systematic forecast failure after shifts in their long-run, or equilibrium, means. Indeed, because they have in-built constant equilibria, their forecasts tend to go up (down) when outcomes go down (up), as they try to converge back to previous equilibria. Consequently, while cointegration captures equilibrium correction, care is required when using such models for genuine out-of-sample forecasts after any forecast failure has been experienced.

Fourth, Castle et al. (2018) have found that selecting a model for forecasting from a general specification that embeds the DGP does not usually entail notable costs compared to using the estimated DGP—an infeasible comparator with non-stationary observational data. Indeed when the exogenous variables need to be forecast, selection can even have smaller MSFEs than using a known DGP. That result matches an earlier finding in Castle et al. (2011) that a selected equation can have a smaller root mean square error (RMSE) for estimated parameters than those from estimating

the DGP when the latter has several parameters that would not be significant on conventional criteria. Castle et al. (2018) suggest using looser than conventional nominal significance levels for in-sample selection, specifically 10% and 16% depending on the number of non-indicator candidate variables, and show that this choice is not greatly affected by whether or not location shifts occur either at, or just after, the forecast origin. The main difficulty is when an irrelevant variable that happens to be highly significant by chance has a location shift, which by definition will not affect the DGP but will shift the forecasts from the model, so forecast failure results. Here rapid updating after the failure will drive that errant coefficient towards zero in methods that minimize squared errors, so will be a transient problem.

Fifth, Castle et al. (2018) also conclude that some forecast combination can be a good strategy for reducing the riskiness of forecasts facing location shifts. Although no known method can protect against a shift after a forecast has been made, averaging forecasts from an econometric model, a robust device and a simple first-order autoregressive model frequently came near the minimum MSFE for a range of forecasting models on 1-step ahead forecasts in their simulation study. This result is consistent with many findings since the original analysis of pooling forecasts in Bates and Granger (1969), and probably reflects the benefits of 'portfolio diversification' known from finance theory. Clements (2017) provides a careful analysis of forecast combination. A caveat emphasized by Hendry and Doornik (2014) is that some pre-selection is useful before averaging to eliminate very bad forecasting devices. For example, the GUM is rarely a good device as it usually contains a number of what transpire to be irrelevant variables, and location shifts in these will lead to poor forecasts. Granger and Jeon (2004) proposed 'thick' modelling as a route to overcoming model uncertainty, where forecasts from all non-rejected specifications are combined. However, Castle (2017) showed that 'thick' modelling by itself neither avoids the problems of model mis-specification, nor handles forecast origin location shifts. Although 'thick' modelling is not formulated as a general-to-simple selection problem, it could be implemented by pooling across all congruent models selected by an approach like *Autometrics*.

References

- Bates, J. M., and Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly*, **20**, 451–468. Reprinted in T. C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.
- Castle, J. L. (2017). Sir Clive W. J. Granger model selection. *European Journal of Pure and Applied Mathematics*, **10**, 133–156. <https://ejpam.com/index.php/ejpam/article/view/2954>.
- Castle, J. L., Clements, M. P., and Hendry, D. F. (2019). *Forecasting: An Essential Introduction*. New Haven, CT: Yale University Press.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3**(1). <https://doi.org/10.2202/1941-1928.1097>.
- Castle, J. L., Doornik, J. A., and Hendry, D. F. (2018). Selecting a model for forecasting. Working paper, Economics Department, Oxford University.
- Castle, J. L., Hendry, D. F., and Martinez, A. B. (2017). Evaluating forecasts, narratives and policy using a test of invariance *Econometrics*, **5**(39). <https://doi.org/10.3390/econometrics5030039>.
- Clements, M. P. (2017). Sir Clive W. J. Granger's Contributions to Forecasting. *European Journal of Pure and Applied Mathematics*, **10**(1), 30–56. <https://www.ejpam.com/index.php/ejpam/article/view/2949>.
- Clements, M. P., and Hendry, D. F. (1996). Intercept corrections and structural change. *Journal of Applied Econometrics*, **11**, 475–494.
- Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, MA: MIT Press.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.
- Granger, C. W. J., and Jeon, Y. (2004). Thick modeling. *Economic Modelling*, **21**, 323–343.
- Hendry, D. F., and Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge, MA: MIT Press.
- Hendry, D. F., and Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, **3**(1). <https://doi.org/10.2202/1941-1928.1100>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

