Forecasting using Economic Systems in Uncertain Environments

Jennifer L. Castle and David F. Hendry

Department of Economics, Oxford University

ESRC Research Methods Festival
3 July 2008
“Trying to predict the future is a mug’s game. But ... we need to have some sort of idea of what the future’s actually going to be like because we are going to have to live there,
“Trying to predict the future is a mug’s game. But ... we need to have some sort of idea of what the future’s actually going to be like because we are going to have to live there, probably next week.”

Douglas Adams
Economic forecasting confronts a non-stationary, evolving world, where model and mechanism differ.

Poor historical track record of econometric systems: forecast failures, and out-performed by ‘naive devices’. Problems date from the early history of econometrics.

Such an adverse outcome is surprising: econometrics uses inter-temporal causal information.

Our aim:
Explain main causes of forecast failure;
Methods to insure against systematic forecast failure;
Some progress towards forecasting during breaks.
Time series of M1 in constant prices
Forecasts of M1 in constant prices

![Graph showing forecasts of M1 in constant prices from 1975 to 1990]
Forecasts of M1 with uncertainty

M1 in constant prices

Research Methods Festival, 3 July 2008 – p.6/43
Outcomes of M1 with uncertainty!
Problem with forecasting is: **future is uncertain.** Forecast uncertainty is intrinsic; but two sources: one we know is present and understand the probabilities; and one due to factors we do not even know exist.

“Because of the things we don’t know [that] we don’t know, the future is largely unpredictable.” Maxine Singer, 1997, *Thoughts of a Nonmillenarian*, p. 39.

In tossing 2 dice, the two sources are:

- the probability any pair of numbers will be face up
- **not knowing that the dice are loaded.**

Second is type of problem in above quote by Singer.
Based on:
individually unpredictable events are ‘regular’ on average.

Economic forecasting similar:
models of economy ‘average’ over possible future ‘shocks’.
Works well for ‘measurable uncertainty’: but—
Economic forecasters confront a difficult environment. Impossible to conceive of all possibilities: economic ‘earthquakes’ seem to occur all too often. **Unmeasured uncertainty important for the future.**

Distributions of events change over time: **non-stationarity.**

Good guides sparse when future is not like past.
AR(1) inflation forecasts

\[
\Delta \hat{\phi} (1970 - 87) \times \pm 2 \hat{\sigma}_{\phi(T)}
\]

\[
\Delta \hat{p}
\]

\[
\Delta p
\]
Stationary scalar first-order autoregressive example:

\[ x_t = \rho x_{t-1} + v_t \quad \text{where} \quad v_t \sim \text{IN} \left[ 0, \sigma_v^2 \right] \quad \text{and} \quad |\rho| < 1. \]

With \( \rho \) known and constant, forecast from \( x_T \) is:

\[ \hat{x}_{T+1|T} = \rho x_T \]

\( D_{x_1^T} (\cdot) \) implies \( D_{x_{T+1}^T} (\cdot) \), producing unbiased forecast:

\[ E \left[ (x_{T+1} - \hat{x}_{T+1|T}) \mid x_T \right] = E \left[ (\rho - \rho) x_T + v_T \right] = 0, \]

with smallest possible variance determined by \( D_{x_1^T} (\cdot) \):

\[ \text{Var} \left[ (x_{T+1} - \hat{x}_{T+1|T}) \right] = \sigma_v^2. \]

Thus:

\[ D_{x_{T+1}^T} (\cdot) = \text{IN} \left[ \rho x_T, \sigma_v^2 \right]. \]

But inflation example illustrates main problem:

Mean has changed over time.
(1) Specification incomplete if (e.g.) vector \( x_t \) not scalar.
(2) Measurement incorrect if (e.g.) observe \( \widetilde{x}_t \) not \( x_t \).
(3) Formulation inadequate if (e.g.) intercept needed.
(4) Modelling wrong if (e.g.) selected \( \rho x_{t-2} \).
(5) Estimating \( \rho \) adds bias, \( (\rho - \mathbb{E}[\hat{\rho}]) x_T \), and variance \( \mathbb{V}[\hat{\rho}] x_T^2 \).
(6) Properties of \( D(v_t) = \mathbb{IN}[0, \sigma_v^2] \) determine \( \mathbb{V}[x_t] \).

(7) Assumed \( v_{T+1} \sim \mathbb{IN}[0, \sigma_v^2] \) but \( \mathbb{V}[v_{T+1}] \) could differ.
(8) Multi-step forecast error \( \sum_{h=1}^{H} \rho^{h-1} v_{T+h} \) has \( \mathbb{V} = \frac{1-\rho^{2H}}{1-\rho^2} \sigma_v^2 \).
(9) If \( \rho = 1 \) have trending forecast variance \( H \sigma_v^2 \).
(10) If \( \rho \) changes could experience forecast failure.

Must be prepared for risks from (1)–(10).
**Estimated AR1:** $\rho = 0.8, T = 40, \sigma^2_{\nu} = 10$
AR1 forecasts: break in $\rho = 0.4$ at $T = 40$

1–step forecasts, known $\rho$ versus estimated $\rho$

5–step forecasts, constant parameters

5–step forecasts, changed $\rho$

5% significance line
Problems hardly disastrous

Small increase in uncertainty from estimating $\rho$; forecast intervals grow quite slowly as $H$ increases.

Little noticeable impact from halving $\rho$ at $T = 40$.

Constancy test hardly rejects false null.

But, slight change to model:

$$x_t = \alpha + \rho x_{t-1} + v_t \quad \text{where} \quad v_t \sim \text{IN} \left[0, \sigma^2_v\right] \quad \text{and} \quad |\rho| < 1.$$  

Everything else the same, except $\alpha = 10$.

Little change in estimation distributions or forecasts: until non-constant $\rho$, for same size and time of break.

Then – catastrophe!!
AR1 forecasts: intercept & break in $\rho$

\[ x_t = 10 + \rho x_{t-1} + v_t \]

Constancy test

1–step forecasts

5–step forecasts
Problems now disastrous

Change due to effect on $E[x_t]$.

In first case $E[x_t] = 0$ before and after shift in $\rho$.

In second: $E[x_t] = \alpha/(1 - \rho)$.

Shifts markedly from 50 to 17.

All models in this class are equilibrium correction:
so fail systematically if $E[\cdot]$ changes.

Huge class of equilibrium-correction models (EqCMS):
regressions; dynamic systems; VARs; DSGEs;
ARCH; GARCH; some other volatility models.

Pervasive and pernicious problem.
Must write conditional expectation as:
\[ \tilde{x}_{T+h|T} = E_{T+h}[x_{T+h}|X_T]. \]

Fine if stationary: \( E_{T+h} = E_T \).
But paradox if \( D_{x_t}(\cdot) \) not constant:
need to know whole future distribution to derive forecast.
**Cannot prove** \( \tilde{x}_{T+h|T} = E_T[x_{T+h}|X_T] \) is useful.

**Empirically-relevant theory needs to allow for:**
model mis-specified for DGP
parameters estimated from inaccurate observations,
on an integrated-cointegrated system,
which intermittently alters unexpectedly from structural breaks.
Possible forecasting problems

Mis-specification, mis-estimation, non-constancy, of deterministic, stochastic, or error components, all could induce forecast failure.

But location shifts are the key problem, namely shifts in parameters of deterministic components. Location shifts easy to detect: see figure 21. Other breaks not so easy to detect: impulse response analyses then unreliable.

Many conventional results change radically when parameter non-constancy: non-causal models can outperform causal; multi-step forecasts more accurate than 1-step; intercept corrections can improve forecasts.
Location shifts in UK unemployment
Forecast failure for UK unemployment
When $f_{t+1}(\cdot) \neq f_t(\cdot)$, forecasting devices robust to location shifts win forecasting competitions.
Location shifts and broken trends

Location shift

Broken trend

Research Methods Festival, 3 July 2008 – p.25/43
**Using $\Delta x_T$ to forecast**

Consider the in-sample DGP:

$$\Delta x_T = \gamma + \alpha \left( \beta' x_{T-1} - \mu \right) + \Psi z_T + v_T, \quad (1)$$

where $z_t$ denotes many omitted effects, with:

$$\Delta x_{T+i} = \gamma^* + \alpha^* \left( (\beta^*)' x_{T+i-1} - \mu^* \right) + \Psi^* z_{T+i} + v_{T+i}. \quad (2)$$

A VEqCM in $x_t$ is used for forecasting:

$$\Delta \hat{x}_{T+i|T+i-1} = \hat{\gamma} + \hat{\alpha} \left( \hat{\beta}' x_{T+i-1} - \hat{\mu} \right). \quad (3)$$

**All main sources of forecast error occur given (2):**

- stochastic and deterministic breaks;
- omitted variables;
- inconsistent parameters;
- estimation uncertainty;
- innovation errors.
Contrast using sequence of $\Delta x_{T+i-1}$ to forecast:

$$\Delta \tilde{x}_{T+i|T+i-1} = \Delta x_{T+i-1}. \quad (4)$$

But because of (2), $\Delta x_{T+i-1}$ is ($i > 1$):

$$\Delta x_{T+i-1} = \gamma^* + \alpha^* ( (\beta^*)' x_{T+i-2} - \mu^* ) + \Psi^* z_{T+i-1} + v_{T+i-1}. \quad (5)$$

Thus, $\Delta x_{T+i-1}$ reflects all the effects needed: parameter changes; no omitted variables; with no estimation issues at all.

**Two drawbacks:**

- $v_{T+i-1}$ doubles innovation error variance;
- variables lagged one extra period – adds ‘noise’. **Trade-off.**

Forecast error is $I(-1)$, so very ‘noisy’, but no systematic failure.
RW and AR(1) forecasts for inflation
Does not refute causal models

Analogy: rocket to moon predicted to land 4th July but oxygen tank exploded and mission was aborted: forecast is systematically and badly wrong.

Outcome not due to bad forecasting models and does not refute Newtonian gravitation theory.

Macro-economic forecast failure occurs regularly. Forecast failure depends on forecast-period events: need not invalidate theory or model, nor be predictable from in-sample tests; neither avoided, nor induced, by congruence.
Objectives:
develop methods for forecasting breaks with robust strategies if breaks incorrectly predicted

First requires that:
(1) breaks are predictable
(2) we have information relevant to that predictability
(3) such information is available at the forecast origin
(4) we have a forecasting model that embodies it
(5) we have a method for selecting that model
(6) resulting forecasts are usefully accurate
Robust strategies

Second builds on considerable recent research:

(7) robust forecasting devices
(8) improved intercept corrections
(9) pooling of forecasts
(10) Also need accurate forecast-error uncertainty measures
Role of information analyzed in *Clements and Hendry* (2005)

New formulation with **two** information sets which potentially might be very different—

one economics: regular forces from agents’ behaviour

other could be politics (say): causes of sudden shifts

No claim that such information actually exists in any given instance, but key to model both if it does

Classic example: one set of forces that lead to outbreak of civil war other factors facilitate its continuation—see (e.g.) *Collier and Hoeffler* (2007)
(2) Relevant information

Depends on which breaks matter

**Location shifts** are most pernicious: induce non-stationarity & systematic forecast failure

Theory in *Clements and Hendry (1998, 2006)*

Other breaks of less relevance for forecasting

So seek information relevant to location shifts: or relevant to ongoing effects as a shift occurs

Forecast failure remains common—and systematic
Several possibilities:

‘leading indicators’—but historical record unimpressive
non-linear functions of variables already in models—same

Rapid information updates at forecast origin—
higher frequency data should help

Forecast-error taxonomy for time disaggregation:
higher frequency does not reduce impacts of breaks
see Castle and Hendry (2008)
But may detect breaks sooner, so adapt better

So consider information outside usual subject matter
Appropriate model form entails non-linear reactions

**Portmanteau test for general form of non-linearity**

*Castle and Hendry (2006)*

Low-dimensional, orthogonalized-representation of polynomial functions

Test only needs $2n$ functions for $n$ linear regressors

Power against up to quintics and inverses thereof

Provides basis for general-to-simple approach: linear model embedded in non-linear general model
(5) Modelling non-linearity

Non-linear model selection—many sub-problems:

(A) specify general form of non-linearity;
   – polynomials, exponentials in orthogonalized regressors;

(B) collinearity between non-linear functions;
   – double demeaning to remove key collinearity;

(C) non-normality: non-linear functions capturing outliers;
   – remove outliers by impulse saturation;

(D) excess numbers of irrelevant variables;
   – super-conservative Gets strategy;

(E) potentially more variables than observations;
   – multi-stage ‘combinatorial selection’;

(F) determine specific form of non-linearity.
   – encompassing tests against specific non-linear forms
     (e.g., ‘ogive’, LSTAR, bilinear, . . .).
(6) Forecast accuracy

Even if location shift is predictable by available information embodied in a well-selected non-linear model problems remain

a] **Breaks alter collinearities** between variables

c] **adverse impact on MSFE if collinearity changes** despite large increase in information content of data

d] **unavoidable**—deleting collinear variables does not help: unless they are actually irrelevant;

e] hence **immediate updating** can be crucial
Simplest conditional regression DGP:

\[ y_t = \beta' z_t + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \text{IN} \left[ 0, \sigma^2_\epsilon \right] \tag{6} \]

with \( z_t \) independent of \( \{ \epsilon_t \} \):

\[ z_t \sim \text{IN}_n \left[ 0, \Sigma \right] \tag{7} \]

for \( \Sigma = H' \Lambda H \) with \( H'H = I_n \).

1-step MSFE for known regressors from (6):

\[ E \left[ \hat{\epsilon}^2_{T+1|T} \right] = \sigma^2_\epsilon \left( 1 + \sum_{i=1}^{n} \frac{\lambda^*_i}{T \lambda_i} \right) \tag{8} \]

where \( E \left[ z_{T+1} z'_{T+1} \right] = \Sigma^* = H' \Lambda^* H \).

If \( \hat{\beta}_{(T)} \) retained, (8) continues to hold. But with updating:

\[ E \left[ \hat{\epsilon}^2_{T+2|T+1} \right] = \sigma^2_\epsilon \left( 1 + \sum_{i=1}^{n} \frac{\lambda^*_i}{T \lambda_i + \lambda^*_i} \right) \tag{9} \]

Reduction depends on smallest eigenvalue ratio.
Impact of breaks in collinearity

1.1
1.2
1.3
1.4

Smallest $\lambda$ ratio = 0.000

$\beta_{\text{Rolling, } k=25}$
$\beta_{\text{Rolling, } k=10}$

In-sample $\beta$
Recursive $\beta$

1.5
1.4
1.3
1.2
1.1

Smallest $\lambda$ ratio = 0.002

$\beta_{\text{Rolling, } k=25}$
$\beta_{\text{Rolling, } k=10}$

1.5
1.4
1.3
1.2
1.1

Smallest $\lambda$ ratio = 0.153

$\beta_{\text{Rolling, } k=25}$
$\beta_{\text{Rolling, } k=10}$

1.5
1.4
1.3
1.2
1.1

Smallest $\lambda$ ratio = 0.67

$\beta_{\text{Rolling, } k=25}$
$\beta_{\text{Rolling, } k=10}$
Use robust forecasting devices
‘Insurance’ after a break to mitigate systematic failure
Hendry (2006): explanation for success of naive devices

(8) Improved intercept corrections
‘Set on track’ at the forecast origin, while smoothing recent corrections: Hendry and Reade (2006)

(9) Pooling of forecasts
‘Model averaging’ can go seriously wrong, but improved by Gets model selection: Hendry and Reade (2004)

(10) Accurate forecast-error uncertainty measures: on-going research
Conclusions

Despite weak assumptions of non-stationary economy, subject to unanticipated structural breaks, model differs from DGP in unknown ways, selected and estimated from unreliable data, can derive many useful insights.


Surprisingly: poor methods; bad models; inaccurate data; and data-based selection not primary causes of systematic mistakes. Main causes are unanticipated large changes affecting forecast period.
Conclusions

Whether breaks are predictable from relevant information available at the forecast origin remains unknown as yet. But progress in developing forecasting models; and methods of testing for and selecting such models.

Predictability theory: 2 information sets, regular and shifts; model latter as non-linear ogive.

Considerable progress since: “The only function of economic forecasting is to make astrology look respectable”
John Kenneth Galbraith


