Forecasting by factors, by variables, by both, or neither?

Jennifer L. Castle*, Michael P. Clements†
and David F. Hendry*

* Institute for New Economic Thinking
at the Oxford Martin School, University of Oxford
† Economics Department, Warwick University

George Washington University, September 2011
Is it better to forecast by factors, by variables, by both, or by neither?

Select forecasting models from candidate set including both regressors and their principal components jointly. Requires selection method to handle perfect collinearity & more regressors than observations.

*Autometrics* allows that, and tackles breaks and outliers by also including an indicator variable for every observation.

Apply to forecasting US GDP and inflation over 1-, 4- and 8-step horizons:

factor models more useful for 1-step ahead; variables tend to be better as the horizon increases; difficult to beat autoregressions.
Three venerable traditions in economic forecasting


Few direct comparisons of first two, none including both. Most comparisons to third,
(1) Setting the scene
(2) Relating variables \( \{z_t\} \) to factors \( \{f_t\} \)
(3) Comparing variable-based and factor-based models
(4) Taxonomy of forecast errors in factor models
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusions
Eight interacting issues affect analyses of forecasting: show answer to our question is context specific:

(i) include both variables & factors in forecasting models;
(ii) role of in-sample model selection in that setting;
(iii) whether breaks are anticipated or not;
(iv) role of more versus less information in forecasting;
(v) type of forecasting model (equilibrium-correction mechanisms–EqCMs);
(vi) data measurement errors near the forecast origin;
(vii) how to evaluate ‘success or failure’ of forecasts;
(viii) nature of DGP itself.

Consider these eight below.
(i) Pooling information

Factor models pool many predictors: or could pool forecasts over many models.

Automatic model selection allows regression to pool information.

Form principal components from \( n \) ‘external’ variables, \( z_t \), where \( n > T \) observations.
Let \( f_t = B z_t \), where \( B \) is matrix of eigenvectors, then \( f_t, \ldots f_{t-s}, z_t, \ldots z_{t-s} \) are possible candidate variables.

Could also include other models’ forecasts as can handle \( n > T \) and perfect collinearity.
(ii) Model selection

Multi-path selection with expanding and contracting block searches can handle perfect collinearity and $n > T$, so not insurmountable to include all $2^{sn}$. *Autometrics* within *PcGive* (see Doornik, 2009) formulates a general unrestricted model (GUM), checking congruence once $N \ll T$ and no perfect collinearity.

Location shifts and outliers removed by impulse-indicator saturation (IIS: see Johansen and Nielsen, 2009, and Castle, Doornik and Hendry, 2011a). 

$\{1\{j=t\}, t = 1, \ldots, T\}$ is complete set of $T$ impulse indicators initial candidates: $f_t, \ldots f_{t-s}, z_t, \ldots z_{t-s}, \{1\{j=t\}, t = 1, \ldots, T\}$

so $N > T$ always occurs when IIS is used.
(iii) Unanticipated location shifts

*Ex ante* forecasting fundamentally different from *ex post* modeling when unanticipated location shifts occur. Can always model shifts after the event, but forecast failure when not anticipated—location shifts most pernicious for forecasting.

Castle, Fawcett and Hendry (2011b) investigate how to forecast breaks, and if fail, how to forecast during breaks: pessimistic due to limited information at location shift occurrence.

Allow for unanticipated location shifts here.
Taxonomy of forecast errors in Clements and Hendry (2005) shows incomplete information *per se* not explanation of forecast failure—unless that information would forecast breaks.

More information cannot lower predictability, omitting explanatory variables leads to inferior forecasts, but large amounts of data will not fix key problem of location shifts, unless can forecast them in advance.

Makridakis and Hibon (2000) conclude parsimonious models best in forecasting competitions, but parsimonious models often robust to location shifts: *jury still out on benefits of more versus less information when forecasting.*
Factor models are often EqCMs: suffer from non-robustness of that class to location shifts. But principles of robust-model formulation from Hendry (2006) apply: any equilibrium-correction system could be differenced prior to forecasting. How a model is used in forecasting matters.
Many ‘solutions’ to forecast failure exacerbate adverse effects of measurement errors near the forecast origin: e.g., differencing doubles their impact.

Averaging mitigates effects of random measurement errors, so factors might reduce such errors.

Models that account for data revisions also a solution.

Frequency of macroeconomic data affects its accuracy, as does nowcasting & ‘real time’ (versus *ex post*) forecasting.

Data measurement errors seem larger in latest data, and during turbulent periods: might favour factor models over approaches that ignore data revisions.
Vast literature on how to evaluate *forecasts*, and using forecasts to evaluate *models*, forecasting *methods*, and economic theory.

Here just report descriptive statistics of forecast performance.
(viii) Nature of the DGP

Nature of the DGP greatly affects forecasting success. Factor model fine if ‘driving forces’ primarily factors, so a few factors account for most variance. Ideal case for factor model forecasting is when DGP is:

\[ x_t = \Upsilon(L) f_t + e_t \]
\[ f_t = \Phi(L) f_{t-1} + \eta_t \]

where \( x_t \) is \( n \times 1 \), \( f_t \) is \( m \times 1 \), \( \Upsilon(L) \) and \( \Phi(L) \) are \( n \times m \) and \( m \times m \), and \( n \gg m \) so low-dimensional \( f_t \) drives co-movements of high-dimensional \( x_t \).

When mean-zero ‘idiosyncratic’ errors \( e_t \) satisfy \( E[e_{i,t}e_{j,t-k}] = 0 \ i \neq j \), and \( E[\eta_te_{t-k}] = 0 \) for all \( k \), then each \( x_{i,t} \) can be optimally forecast using \( f_t \) and lags of \( x_{i,t} \).
Let $\lambda_i(L)'$ denote the $i^{th}$ row of $\Upsilon(L)$, then ignoring parameter estimation uncertainty:

$$
\mathbb{E}_t[\mathbf{x}_{i,t+1} | \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots] \\
= \mathbb{E}_t[\lambda_i(L)' \mathbf{f}_{t+1} | \mathbf{f}_t, \mathbf{f}_{t-1}, \ldots] + \mathbb{E}_t[e_{i,t+1} | e_{i,t}, e_{i,t-1}, \ldots] \\
= \alpha(L)' \mathbf{f}_t + \delta(L) x_{i,t}
$$

Absent structural breaks, delivers best forecasts. But factor structures may not be good representation of macroeconomy: driving forces may be variables, but factor approximations may parsimoniously capture linear combinations of effects.
Selection with $N$ (variables) $> T$ (observations) seems to hinge on sparsity of parameter vector, $\beta$, in local DGP:

- must have fewer non-zero **functions** of parameters than $T$, otherwise LDGP is not identifiable.

**Factor models are ‘solution’ to dimension constraints.** When selecting with both principal components and variables, sparsity is **not** required: $\beta$ could contain all non-zero elements, yet fewer regressors than $T$.

Automatic selection handles $N \gg T$ & perfect collinearity, so no binding dimensionality constraint: **include all variables & principal components** to retain both key individual determinants and factors.
Our approach

Linear models selected in-sample with IIS from:
(a) a large set of variables,
(b) over those variables’ principal components (PCs), and
(c) over a candidate set including both.
Initial model has $N > T$,
and in case (c), is perfectly collinear.
(1) **Setting the scene**

(2) **Relating variables** $\{z_t\}$ **to factors** $\{f_t\}$

(3) Comparing variable-based and factor-based models

(4) Taxonomy of forecast errors in factor models

(5) Forecast failure in equilibrium-correction models

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

**Conclusions**
Consider $n < T$ weakly stationary $\{z_t\}$ over $t = 1, \ldots, T$ generated by VAR(1) with intercept $\pi$:

$$z_t = \pi + \Pi z_{t-1} + v_t \text{ where } v_t \sim \mathcal{IN}_n [0, \Omega_v]$$  \hspace{1cm} (1)

all eigenvalues of $\Pi$ inside unit circle:

$$E[z_t] = \pi + \Pi E[z_{t-1}] = \pi + \Pi \mu = \mu$$

where $\mu = (I_n - \Pi)^{-1}\pi$.

Principal-component description is:

$$z_t = \Psi f_t + e_t \text{ with } e_t \sim \mathcal{ID}_n [0, \Omega_e]$$  \hspace{1cm} (2)

$f_t \sim \mathcal{ID}_m[\kappa, P]$ of dimension $m \leq n$, $\Psi$ is $n \times m$, $E[f_t e_t'] = 0$.

When $E[f_t] = \kappa$ and $E[e_t] = 0$, from (2):

$$E[z_t] = \Psi E[f_t] + E[e_t] = \Psi \kappa = \mu$$  \hspace{1cm} (3)
Eigenvector decompositions

Then:

\[
E \left[ (z_t - \mu) (z_t - \mu)' \right] = \Psi E \left[ (f_t - \kappa) (f_t - \kappa)' \right] \Psi' + E \left[ e_t e_t' \right]
\]

\[
= \Psi P \Psi' + \Omega_e = M
\]  \hspace{1cm} (4)

where \( P \) is \( m \times m \) diagonal and hence \( z_t \sim D_n[\mu, M] \). Let:

\[
M = B\Lambda B'
\]  \hspace{1cm} (5)

where \( B'B = I_n \), so \( B^{-1} = B' \) with eigenvalues ordered from largest downwards:

\[
B' = \begin{pmatrix} B'_1 \\ B'_2 \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix},
\]  \hspace{1cm} (6)

where \( \Lambda_{11} \) is \( m \times m \), with \( B'_1 MB_1 = \Lambda_{11} \) and:

\[
B\Lambda B' = B'_1 \Lambda_{11} B'_1 + B'_2 \Lambda_{22} B'_2.
\]
Consequently, from (2) and (6):

\[ B'(z_t - \mu) = B'(\Psi(f_t - \kappa) + e_t) = f_t - \kappa \]  \hspace{1cm} (7)

If only \( m \) actually matter, \( B'_1 \) weights the \( z_t \) to produce \( f_{1,t} \) where:

\[ B'_1(z_t - \mu) = f_{1,t} - \kappa_1 \]  \hspace{1cm} (8)

where only \( f_{1,t} \) are relevant to explaining \( y_t \).
When no factors matter, DGP over \( t = 1, \ldots, T \) is:
\[
y_t = \beta' z_t + \epsilon_t
\] (9)
where \( \epsilon_t \sim \text{IN} \left[ 0, \sigma^2 \right] \). Regressors generated by:
\[
z_t \sim \text{IN}_n \left[ 0, \sigma^2 \Omega \right], \quad t = 1, \ldots, T
\] (10)
where \( \Omega = (1 - \rho)I_n + \rho \iota\iota' \), \( T = 100 \), with \( \rho = 0.5, 0.9 \), \( n = 2, 10, 50 \) using \( M = 10,000 \) replications in Autometrics at \( \alpha = 0.05, 0.01 \), with no diagnostic checking.

Most general GUM is given by:
\[
y_t = \gamma' z_t + \delta' f_t + \nu_t
\] (11)
where \( \delta = 0 \), leading to \( 2n \) regressors, with \( n \) collinearities.

Table 1 records the gauge: close to \( \alpha \) when no collinearity, well under with collinearity, till \( n = 50 \), when expanding and contracting block searches return \( g \simeq \alpha \).
Simulation outcomes under the null

<table>
<thead>
<tr>
<th></th>
<th>$\alpha = 0.05$</th>
<th></th>
<th>$\alpha = 0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 2$</td>
<td>$n = 10$</td>
<td>$n = 50$</td>
</tr>
<tr>
<td>$\rho = 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>0.052</td>
<td>0.048</td>
<td>0.031</td>
</tr>
<tr>
<td>(ii)</td>
<td>0.052</td>
<td>0.048</td>
<td>0.031</td>
</tr>
<tr>
<td>(iii)</td>
<td>0.024</td>
<td>0.002</td>
<td>0.047</td>
</tr>
<tr>
<td>$\rho = 0.9$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>0.052</td>
<td>0.057</td>
<td>0.031</td>
</tr>
<tr>
<td>(ii)</td>
<td>0.052</td>
<td>0.057</td>
<td>0.031</td>
</tr>
<tr>
<td>(iii)</td>
<td>0.026</td>
<td>0.002</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Table 1: Gauge for selection with (i) just variables, (ii) just factors, and (iii) variables and factors.
First, close to a ‘factor DGP’: factor structure of $z_t$ in (10) matches relation between $y_t$ and $z_t$ in (9). For $n = 3$:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

(12)

which has an eigenvector of $h_1 = \iota = (1 \ldots 1)'$ so:

$$\phi f_{1,t} = \phi' z_t = \beta' z_t.$$ 

(13)

When DGP in (9) has $\phi = \iota' \beta$:

$$y_t = \phi f_{1,t} + \epsilon_t$$

(14)

with non-centrality $\psi$ of the $t$-test of $\phi = 0$:

$$\psi = \frac{\hat{\phi}}{\text{SE} [\hat{\phi}]} \simeq \frac{\phi \sigma_f \sqrt{n (1 + (n - 1) \rho)}}{\sigma_\epsilon}$$

(15)
When $\sigma_f^2 = \sigma_\epsilon^2 = 1$, $\rho = 0.9$ and $n = 3$, ratio of factor to variable non-centralities is 7.5. (Ratios of 28.8 at $n = 10$ and 162 at $n = 50$)

When many small effects, in this setting PCs retained with high probability relative to retention of individual regressors.

Huge advantage to representing individually-insignificant effects in $z_t$ by factors: large reduction in RMSEs relative to just estimating DGP. Bias correction downweights retained irrelevant factors, yielding yet smaller RMSEs; first principal component highly significant, captures most of variation in $y$
After selection, denote retained $\mathbf{z}_t$ and $\mathbf{w}_t$ by $\mathbf{z}_{r,t}$ and $\mathbf{w}_{s,t}$, with coefficients $\tilde{\gamma}_r$ and $\tilde{\delta}_s$:

$$y_t = \tilde{\gamma}'_r \mathbf{z}_{r,t} + \tilde{\delta}'_s \mathbf{w}_{s,t} + v_t$$  \hspace{1cm} (16)

From (37) the retained principal components are:

$$\mathbf{w}_{s,t} = \hat{\mathbf{H}}_s \tilde{\mathbf{z}}_t$$

Solving out for the principal components results in:

$$y_t = \tilde{\gamma}'_r \mathbf{z}_{r,t} + \tilde{\delta}'_s \mathbf{w}_{s,t} + v_t$$

$$= \tilde{\gamma}'_r \mathbf{z}_t + \tilde{\delta}'_s \mathbf{w}_t + v_t$$

$$= \lambda_s + \tilde{\gamma}'_s \mathbf{z}_t + v_t$$  \hspace{1cm} (17)

$\gamma_{r*}$ and $\delta_{s*}$ augment $\gamma_r$ and $\delta_s$ by zeros for non-selected elements.
Evaluating selection

First $k$ regressors relevant, $n - k$ irrelevant. $	ilde{\beta}_{j,i}^*$ denotes OLS coefficient estimate after selection for $j$th regressor in replication $i$, with $M$ replications.

$$\bar{\beta}_j^* = \frac{\sum_{i=1}^{M} \left[ \tilde{\beta}_{j,i}^* 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right) \right]}{\sum_{i=1}^{M} 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right)}$$

$$\text{bias}_j = \frac{\sum_{i=1}^{M} \left[ (\tilde{\beta}_{j,i}^* - \beta_j) 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right) \right]}{\sum_{i=1}^{M} 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right)}$$

$$\text{RMSE}_j = \left[ \frac{\sum_{i=1}^{M} (\tilde{\beta}_{j,i}^* - \beta_j)^2 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right)}{\sum_{i=1}^{M} 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right)} \right]^{1/2}$$

or $\beta_j^2$ when $\sum_{i=1}^{M} 1 \left( \tilde{\beta}_{j,i}^* \neq 0 \right) = 0$. 

Castle, Clements & Hendry
Forecasting by factors, by variables, by both, or neither? – p.26/76
Correctly approximating small effects by PCs

<table>
<thead>
<tr>
<th></th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>[B]</th>
<th>[C]</th>
<th>[D]</th>
<th>[E]</th>
<th>[F]</th>
<th>[E]</th>
<th>[F]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>5%</td>
<td>1%</td>
<td>1%</td>
<td>5%</td>
<td>5%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 10; \psi_i = 1; \rho = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.998</td>
<td>0.995</td>
<td>0.995</td>
<td>1.012</td>
<td>1.012</td>
<td>1.004</td>
<td>0.993</td>
<td>1.001</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>$\bar{n}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.022</td>
<td>23.54</td>
<td>17.13</td>
<td>31.56</td>
<td>25.50</td>
<td>-0.023</td>
<td>-0.069</td>
<td>-0.075</td>
<td>-0.082</td>
<td>-0.083</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.143</td>
<td>0.253</td>
<td>0.217</td>
<td>0.330</td>
<td>0.290</td>
<td>0.144</td>
<td>0.076</td>
<td>0.052</td>
<td>0.044</td>
<td>0.032</td>
</tr>
<tr>
<td>$n = 10; \psi_i = 1; \rho = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.998</td>
<td>0.994</td>
<td>0.994</td>
<td>1.010</td>
<td>1.010</td>
<td>1.046</td>
<td>1.033</td>
<td>1.039</td>
<td>1.039</td>
<td></td>
</tr>
<tr>
<td>$\bar{n}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>-0.018</td>
<td>53.15</td>
<td>40.26</td>
<td>71.86</td>
<td>61.05</td>
<td>-0.018</td>
<td>-0.023</td>
<td>-0.023</td>
<td>-0.024</td>
<td>-0.024</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.318</td>
<td>0.574</td>
<td>0.503</td>
<td>0.752</td>
<td>0.682</td>
<td>0.320</td>
<td>0.157</td>
<td>0.102</td>
<td>0.083</td>
<td>0.056</td>
</tr>
</tbody>
</table>

If link of $y_t$ to $z_t$ in (9) does not match correlations in (10), (e.g., $\beta_i$ alternate as $\pm \phi$) need all 3 factors. Other two eigenvectors of (12) are:

$$h_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad h_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

so $\beta'z_t = \phi \begin{pmatrix} 1 & -1 & 1 \end{pmatrix}z_t$ equals $h_1 + 2h_2 - 4h_3$.

$\beta_j = -a \times -1^j$, for $j = 1, \ldots, n$ where $a$ calibrated for $|t| = 1$

A final case is where a subset of the $z_t$ are irrelevant so $\beta' = \phi \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$ or $\phi \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}$. The former, e.g., is $\phi (2h_1 + h_2 + h_3)$ (again needing all three factors), the latter is $\phi h_2$. 
Simulating poor small-effects approximation by PCs

<table>
<thead>
<tr>
<th></th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>[D]</th>
<th>[E]</th>
<th>[F]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>$n = 10; \psi_i = (-1)^i; \rho = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.998</td>
<td>1.021</td>
<td>1.021</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>$\bar{\eta}$</td>
<td>0.636</td>
<td>0.636</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias (+)</td>
<td>-0.020</td>
<td>22.02</td>
<td>12.63</td>
<td>-0.016</td>
<td>-5.470</td>
<td>-7.411</td>
</tr>
<tr>
<td>Bias (−)</td>
<td>-0.025</td>
<td>-21.92</td>
<td>-12.56</td>
<td>-0.029</td>
<td>5.464</td>
<td>7.406</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.143</td>
<td>0.243</td>
<td>0.182</td>
<td>0.144</td>
<td>0.145</td>
<td>0.130</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>[A]</th>
<th>[B]</th>
<th>[C]</th>
<th>[D]</th>
<th>[E]</th>
<th>[F]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>$n = 10; \psi_i = (-1)^i; \rho = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.998</td>
<td>1.016</td>
<td>1.016</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$\bar{\eta}$</td>
<td>0.650</td>
<td>0.650</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias (+)</td>
<td>0.011</td>
<td>46.83</td>
<td>33.21</td>
<td>0.019</td>
<td>-12.38</td>
<td>-16.78</td>
</tr>
<tr>
<td>Bias (−)</td>
<td>-0.047</td>
<td>-47.09</td>
<td>-33.37</td>
<td>-0.056</td>
<td>12.38</td>
<td>16.78</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.318</td>
<td>0.522</td>
<td>0.420</td>
<td>0.320</td>
<td>0.325</td>
<td>0.289</td>
</tr>
</tbody>
</table>

[A] estimating DGP;  
[B] selection from DGP by Autometrics;  
[C] bias correction in [B];  
[D] estimation of factor model;  
[E] 1-cut selection from factor model;  
[F] bias correction of [E].
Setting the scene

Relating variables \( \{z_t\} \) to factors \( \{f_t\} \)

Comparing variable-based and factor-based models

Taxonomy of forecast errors in factor models

Forecast failure in equilibrium-correction models

Model selection with both factors and variables

Forecasting US GDP and inflation

Conclusions
Variable-based and factor-based models

The in-sample DGP for $y_t$ is:

$$y_t = \beta_0 + \beta' z_{t-1} + \rho y_{t-1} + \epsilon_t$$  \hspace{1cm} (18)

where $|\rho| < 1$ and $\epsilon_t \sim \text{IN}[0, \sigma^2_\epsilon]$, so:

$$E[y_t] = \beta_0 + \beta' E[z_{t-1}] + \rho E[y_{t-1}] = \beta_0 + \beta' \mu + \rho \delta = \delta$$  \hspace{1cm} (19)

where $\delta = (\beta_0 + \beta' \mu)/(1 - \rho)$ so (18) is the EqCM:

$$y_t - \delta = \beta' (z_{t-1} - \mu) + \rho (y_{t-1} - \delta) + \epsilon_t$$  \hspace{1cm} (20)

Only the subset $z_{a,t-1}$ matter substantively:

$$y_t - \delta \approx \beta'_a (z_{a,t-1} - \mu_a) + \rho_a (y_{t-1} - \delta) + u_t$$  \hspace{1cm} (21)

Alternatively, combinations $f_{1,t}$ might matter:

$$y_t - \delta = \beta' f_{t-1} - \kappa) + \rho (y_{t-1} - \delta) + \epsilon_t$$  \hspace{1cm} (22)

so that:

$$y_t - \delta = \tau'_1 (f_{1,t-1} - \kappa_1) + \rho_1 (y_{t-1} - \delta) + w_t$$  \hspace{1cm} (23)
When both variables and PCs are allowed:

\[ y_t - \delta = \beta'_b (z_{b,t-1} - \mu_b) + \tau'_2 (f_{2,t-1} - \kappa_2) + \rho_b (y_{t-1} - \delta) + \epsilon_t \]  \hspace{1cm} (24)

Use IIS to handle location shifts and outliers in-sample.

If 2n variables and factors at significance level \( \alpha \), 2\( \alpha n \) will be adventitiously significant on average—\textbf{one} retained by chance when \( n = 100 \) and \( \alpha = 0.005 \):

- critical value of \( c_\alpha \approx 3.10 \).

When DGP for \( y_t \) is an AR(1):

\[ y_t = \gamma_0 + \gamma_1 y_{t-1} + \nu_t \]

If search \textbf{only} over factors, retain many \( f_t \) to approximate \( y_{t-1} \) so starting model always includes \( y_{t-1} \).
Forecast \( \{y_{T+h}\} \) over horizon \( h = 1, \ldots, H \), from forecast origin at \( T \), with information set \( Z^1_T = (z_1 \ldots z_T) \) and \( (y_1 \ldots y_T) \).

For in-sample estimates \( \{\hat{f}_t\} \), 1-step forecasts are:

\[
\hat{y}_{T+1|T} = \hat{\delta} + \hat{\beta}'_b (z_{b,T} - \hat{\mu}_b) + \hat{\tau}'_2 (\hat{f}_{2,T} - \hat{\kappa}_2) + \hat{\rho} (\hat{y}_T - \hat{\delta})
\]  
(25)

where \( \hat{y}_T \) is the ‘flash’ estimate.

Use multi-step estimation for \( h \)-step ahead forecasts:

\[
\hat{y}_{T+h|T} = \hat{\delta}_{(h)} + \hat{\beta}'_{b,(h)} (z_{b,T} - \hat{\mu}_b) + \hat{\tau}'_{2,(h)} (\hat{f}_{2,T} - \hat{\kappa}_2) + \hat{\rho}_{(h)} (\hat{y}_T - \hat{\delta}_{(h)})
\]  
(26)

so \( \hat{u}_{T+h|T} = y_{T+h} - \hat{y}_{T+h|T} \), a moving average of order \( (h - 1) \).
(1) Setting the scene
(2) Relating variables \( \{z_t\} \) to factors \( \{f_t\} \)
(3) Comparing variable-based and factor-based models
(4) **Taxonomy of forecast errors in factor models**
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusions
Location shifts

DGP depends on $z_{t-1}$ and $y_{t-1}$, forecasting model incorporates a subset of the factors.

Highlight principle sources of forecast bias and forecast-error variance after location shifts where DGP changes at $T$ to:

$$y_{T+h} = \delta^* + \beta' (z_{T+h-1} - \mu^*) + \rho (y_{T+h-1} - \delta^*) + \epsilon_{T+h}$$  \hspace{1cm} (27)

for $h = 1, \ldots, H$.

Mapping to principal components yields:

$$y_{T+h} = \delta^* + \tau' (f_{T+h-1} - \kappa^*) + \rho (y_{T+h-1} - \delta^*) + \epsilon_{T+h}$$  \hspace{1cm} (28)

where for now $\tau$ and $\rho$ remain at in-sample values during forecast period.
1-step forecast-error taxonomy distinguishes 11 sources.

From (28) minus (25), for $h = 1$:

$$\hat{u}_{T+1|T} = \left( \delta^* - \hat{\delta} \right) + \tau' (f_T - \kappa^*) - \hat{\tau}'_1 \left( \hat{f}_{1,T} - \hat{\kappa}_1 \right)$$

$$+ \rho (y_T - \delta^*) - \hat{\rho} \left( \hat{y}_T - \hat{\delta} \right) + \epsilon_{T+1}.$$

Using $\tau'_1 (\kappa^*_1 - \kappa_1) + \tau'_2 (\kappa^*_2 - \kappa_2) = \tau' \left( \kappa^* - \kappa \right)$, yields table 2.
**Table 2: \( \hat{u}_{T+1|T} = \ldots \)**

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>((1 - \rho) (\delta^* - \delta))</td>
<td>[A] equilibrium-mean shift</td>
</tr>
<tr>
<td>(-\tau' (\kappa^* - \kappa))</td>
<td>[B] factor-mean shift</td>
</tr>
<tr>
<td>(+ (1 - \rho) (\delta - \delta))</td>
<td>[C] equilibrium-mean estimation</td>
</tr>
<tr>
<td>(-\tau'_1 (\kappa_1 - \hat{\kappa}_1))</td>
<td>[D] factor-mean estimation</td>
</tr>
<tr>
<td>(+ \rho (y_T - \hat{y}_T))</td>
<td>[E] flash estimate error</td>
</tr>
<tr>
<td>(+ \tau'<em>1 (f</em>{1,T} - \hat{f}_{1,T}))</td>
<td>[F] factor estimate error</td>
</tr>
<tr>
<td>(+ \tau'<em>2 (f</em>{2,T} - \kappa_2))</td>
<td>[G] factor approximation error</td>
</tr>
<tr>
<td>(+ (\tau_1 - \hat{\tau}<em>1)' (\hat{f}</em>{1,T} - \kappa_1))</td>
<td>[H] factor estimation covariance</td>
</tr>
<tr>
<td>(+ (\rho - \hat{\rho}) (\hat{y}_T - \delta))</td>
<td>[I] flash estimation covariance</td>
</tr>
<tr>
<td>(+ (\tau_1 - \hat{\tau}_1)' (\kappa_1 - \hat{\kappa}_1))</td>
<td>[J] parameter estimation covariance</td>
</tr>
<tr>
<td>(+ \epsilon_{T+1})</td>
<td>[K] innovation error</td>
</tr>
</tbody>
</table>
Interpretation

If unbiased estimates, neglecting terms of \( O_p(T^{-1}) \):

\[
E \left[ \hat{u}_{T+1|T} \right] \simeq (1 - \rho) \left( \delta^* - \delta \right) - \tau' (\kappa^* - \kappa) \\
+ \rho (y_T - E[\hat{y}_T]) + \tau'_1 (f_{1,T} - E[\hat{f}_{1,T}])
\]

(29)

[A] and [B] primary determinants of forecast bias: data and factor-estimation errors [E] & [F] contribute. Those & remaining terms add to forecast-error variance. Factor approximation error not in (29) as \( E[f_{2,T}] = \kappa_2 \).

Even if [E] and [F] negligible, shifts could be large, if in (1):

\[
\pi^* = \pi + 1_{(t \geq T)} \theta \quad \text{for } h = 1, \ldots, H
\]

(30)

so permanently shifts at \( T \):

\[
\pi = (I_n - \Pi) \Psi \kappa
\]

When \( \Pi \) and \( \Psi \) are constant, \( \kappa \) will shift, so for \( n = m \):

\[
\kappa^* = \Psi^{-1} (I_n - \Pi)^{-1} \pi^* = \kappa + 1_{(t \geq T)} \Psi^{-1} (I_n - \Pi)^{-1} \theta
\]

(31)
Forecast-error biases entailed by equilibrium-mean shifts within model of $y_{T+1}$ ($\delta^* \neq \delta$) or in external variables ($\kappa^* \neq \kappa$) irrespective of inclusion or exclusion of associated factors: approximation error does not induce such a problem.

If $\tau$ and $\rho$ change in the forecast period:

$$y_{T+1} = \delta^* + \tau^* (f_T - \kappa^*) + \rho^* (y_T - \delta^*) + \epsilon_{T+1} \quad (32)$$

the change in $\tau$ induces an additional error term:

$$\tau^* (f_T - \kappa^*) - \tau' (f_T - \kappa^*) = (\tau^* - \tau') (f_T - \kappa^*)$$

interacting with the location shift, but otherwise is benign—this additional term will not contribute to bias when $\kappa^* = \kappa$.

Similarly, the shift in $\rho$ only affects forecast bias if $\delta$ also changes via:

$$(\rho^* - \rho) (y_T - \delta^*)$$

which has a zero expectation when $\delta^* = \delta$. 
Principal sources of forecast error for an AR(1) model ignoring parameter estimation uncertainty and data mis-measurement:

\[ y_t = \delta + \gamma (y_{t-1} - \delta) + v_t \]  

(33)

when the forecast period DGP is (28). Omitting factors will alter \( \gamma \), so \( \gamma \neq \rho \), but not affect \( \delta \). Forecast error from the AR(1) model is \( \hat{v}_{T+1|T} \), so:

\[
\hat{v}_{T+1|T} = (1 - \rho) (\delta^* - \delta) - \tau' (\kappa^* - f_T) + (\rho - \gamma) (y_T - \delta)
\]

with a forecast bias of:

\[
E \left[ \hat{v}_{T+1|T} \right] = (1 - \rho) (\delta^* - \delta) - \tau' (\kappa^* - \kappa)
\]

matching the two leading terms in (29) for the bias of the factor-forecasting model.

Including ‘correct’ factors, a subset, or none has no effect on forecast biases following a location shift.
(1) Setting the scene

(2) Relating variables $\{z_t\}$ to factors $\{f_t\}$

(3) Comparing variable-based and factor-based models

(4) Taxonomy of forecast errors in factor models

(5) **Forecast failure in equilibrium-correction models**

(6) Model selection with both factors and variables

(7) Forecasting US GDP and inflation

Conclusions
The equilibrium-correction problem

Forecasting is not a one-off venture, so compare models as the origin moves through time. All models will fail when there is an unknown location shift but speed and extent to which forecasts recover differs. ‘Equilibrium-correction’ models, to which (25) belongs, lack adaptability over time.

Write (25) for 1-step forecasts:

\[ \Delta \hat{y}_{T+1|T} = \hat{\tau}_1' (\hat{f}_{1,T} - \hat{\kappa}_1) + (\hat{\rho} - 1) (\hat{y}_T - \hat{\delta}) \]

so that \( E[\Delta \hat{y}_{T+1|T}] \approx 0 \), whereas the DGP is:

\[ \Delta y_{T+1} = \tau' (f_T - \kappa^*) + (\rho - 1) (y_T - \delta^*) + \epsilon_{T+1} \]  

(34) expected value is non-zero when there are locations shifts:

\[ E[\Delta y_{T+1}] = \tau' E[f_T - \kappa^*] + (\rho - 1) E[y_T - \delta^*] \]

\[ = \tau'_1 (\kappa_1 - \kappa_1^*) + (\rho - 1) (\delta - \delta^*) \]  

(35)
Correcting equilibrium corrections

In EqCMs, problem persists: forecasting $T + 2$ from $T + 1$ even for known in-sample parameters, etc:

$$
\Delta \hat{u}_{T+2|T+1} = \Delta y_{T+2} - \Delta \hat{y}_{T+2|T+1}
$$

$$
= \tau_1' (\kappa_1 - \kappa^*_1) + (\rho - 1) (\delta - \delta^*) + \epsilon_{T+1}
$$

Use a robust forecasting device exploiting (35):

$$
\Delta \tilde{y}_{T+2|T+1} = \Delta y_{T+1} + \tilde{\tau}_1' \Delta \hat{f}_{1,T} + (\bar{\rho} - 1) \Delta \hat{y}_T
$$

Again under simplifying assumptions, forecast error

$$
\Delta \tilde{u}_{T+2|T+1} = \Delta y_{T+2} - \Delta \tilde{y}_{T+2|T+1}
$$

is:

$$
\Delta \tilde{u}_{T+2|T+1} = \tau_1' (f_{T+1} - \kappa^*) + (\rho - 1) (y_{T+1} - \delta^*) + \epsilon_{T+2}
$$

$$
- \tau_1' (f_T - \kappa^*) - (\rho - 1) (y_T - \delta^*) - \epsilon_{T+1}
$$

$$
= \tau_1' \Delta f_{1,T+1} + (\rho - 1) \Delta y_{T+1} + \Delta \epsilon_{T+2}
$$

(36)

so less dependent on location shifts.
Illustrate by forecasting Japanese exports over 2008(7)–2011(6)

Model selected at 1% by Autometrics, 2000(1)–2008(6):

\[
\hat{y}_t = 0.68y_{t-1} + 0.26y_{t-3} + 0.121_{2000(2)} + 0.121_{2002(1)}
\]

\[
\hat{\sigma} = 0.039 \quad \chi^2(2) = 0.65
\]

\[
F_{het}(4, 95) = 1.02 \quad F_{ar}(6, 92) = 0.97 \quad F_{reset}(2, 96) = 2.1
\]

\[1_{200z(x)} \text{ indicators for } 200z(x).\]

Almost ‘robust’ by near second unit root.

The corresponding robust device, therefore, was:

\[
\tilde{y}_t = y_{t-1} + 0.95\Delta y_{t-1} \quad \tilde{\sigma} = 0.0755
\]

Their respective forecasts follow.
Model-based forecasts 2008(7)–2011(6)

Castle, Clements & Hendry

Forecasting by factors, by variables, by both, or neither? – p.45/76
Robust forecasts 2008(7)–2011(6)

Forecasting by factors, by variables, by both, or neither? – p.46/76
Typical pattern: robust device avoids forecast failure, at an insurance cost when no shifts occur. Overall RMSFEs of 0.124 versus 0.098.
When factor models are EqCMs, location shifts:
(a) affect the mapping between the $z_{T+1}$ and $f_{T+1}$ by changes in the weights:
Stock and Watson (2009) find a relatively innocuous effect;
(b) breaks in coefficients of zero-mean variables or factors not problematic;
(c) location shifts induce systematic mis-forecasting.
In their forecasting exercise, variables differenced once, so large equilibrium-mean shifts unlikely:
partially robust to previous location, but not growth, shifts.
If the differenced-data version of (33) is used, forecasting $T + 2$ from $T + 1$:

$$\Delta \tilde{y}_{T+2} | T+1 = \gamma \Delta y_{T+1}$$

when:

$$\Delta y_{T+2} = \tau^* \Delta f_{T+1} + \rho^* \Delta y_{T+1} + \Delta \epsilon_{T+2}$$

we have:

$$\tilde{v}_{T+2} | T+1 = \tau^* \Delta f_{T+1} + (\rho^* - \gamma) \Delta y_{T+1} + \Delta \epsilon_{T+2}$$

which is close to (36).
(1) Setting the scene
(2) Relating variables $\{z_t\}$ to factors $\{f_t\}$
(3) Comparing variable-based and factor-based models
(4) Taxonomy of forecast errors in factor models
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusions
Automatic model selection

Selection gives zero weight to insignificant variables, PCs place small non-zero weight even if variables have low correlation with other $z$s.

Selection avoids separate choice of PCs–based on covariation between them and $y$: does not require correlation structure between $z$s and $y$ to be similar to within $z$s.

Empirically, retained PCs tend not to be first few, so correlation structure differs for $y$ and disaggregates.

Multi-path search allows perfectly collinear candidate regressors and $N > T$.

*Autometrics* uses expanding and contracting block searches: ‘sieve’ continues until $N < T$ and no singularities. Tree search then applied.
(1) Setting the scene
(2) Relating variables \( \{z_t\} \) to factors \( \{f_t\} \)
(3) Comparing variable-based and factor-based models
(4) Taxonomy of forecast errors in factor models
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusions
Based on Stock and Watson (2009). Forecast quarterly US GDP growth, change in inflation & level forecasts for GDP and inflation: $H = 1997q1–2006q4$. 144 seasonally-adjusted observations, estimation and selection over $T = 1962q3–2006q4$, after lags. $n = 109$ disaggregates are candidate $z$s and used for $f$s. Original data in logs first or second differenced Sample principal components computed as:

$$\hat{F} = \hat{B}'\tilde{Z}$$

(37)

where $\hat{B}$ is matrix of eigenvectors and $\tilde{Z}$ is standardized data.
Stock and Watson (2009) identify a break in 1984: coefficients of factors in forecasting models not constant across this period.

Factors well estimated by PCs even when loadings unstable: more accurate factor-model forecasts result from whole-sample factors but forecasting models based only on the period after 1984.

Use full-sample PCs and full-sample selection with IIS: allows for multiple breaks and outliers jointly with selection, little efficiency loss under null of no breaks.
Forecasting models selected from GUM:

\[ \Delta y_t = \gamma_0 + \sum_{j=J_a}^{J_b} \gamma_j \Delta y_{t-j} + \sum_{i=1}^{n} \sum_{j=J_a}^{J_b} \beta_{i,j} \Delta z_{i,t-j} + \sum_{k=1}^{n} \sum_{j=J_a}^{J_b} \phi_{k,j} f_{k,t-j} + \sum_{l=1}^{T} \delta_l \{l=t\} + \epsilon_t \]  

(i) \( \phi = 0 \), i.e. select over variables only;
(ii) \( \beta = 0 \), select over factors only; and
(iii) \( \phi \neq 0 \) and \( \beta \neq 0 \), jointly selecting variables and factors;

(a) \( \delta = 0 \), no IIS; and
(b) \( \delta \neq 0 \), with IIS,

1-step: \( J_a = 1, J_b = 4 \);
4-step: \( J_a = 4, J_b = 7 \);
8-step: \( J_a = 8, J_b = 11 \).
Choice of significance level

3 benchmark forecasts:

\[ \Delta \hat{y}_{T+k+h}^{RW} = \Delta y_{T+k} \]
\[ \Delta \hat{y}_{T+k+h}^{AR(D)} = \hat{\beta}_0 + \hat{\beta}_1 \Delta y_{T+k} \]
\[ \Delta \hat{y}_{T+k+h}^{AR(I)} = \sum_{i=1}^{h-1} \hat{\gamma}_0 \hat{\gamma}_1^i + \hat{\gamma}_1^h \Delta y_{T+k} \]

for \( k = 0, \ldots, H - h \) and \( h = 1, 4 \) and 8.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factors</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>No IIS</td>
<td>IIS</td>
<td>No IIS</td>
</tr>
<tr>
<td>( N )</td>
<td>441</td>
<td>579</td>
</tr>
<tr>
<td>( \alpha %)</td>
<td>1 (4.4)</td>
<td>0.5 (3)</td>
</tr>
<tr>
<td>( \alpha %)</td>
<td>0.5 (2.2)</td>
<td>0.1 (0.6)</td>
</tr>
</tbody>
</table>

Significance levels used in model selection (%); expected null retention numbers in parentheses. \( T = 138 \).
In sample fit: GDP growth with IIS

<table>
<thead>
<tr>
<th></th>
<th>1-step</th>
<th></th>
<th>4-step</th>
<th></th>
<th>8-step</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cons</td>
<td>Super</td>
<td>Cons</td>
<td>Super</td>
<td>Cons</td>
<td>Super</td>
</tr>
<tr>
<td>Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>0.559</td>
<td>0.669</td>
<td>0.629</td>
<td>0.711</td>
<td>0.702</td>
<td>0.799</td>
</tr>
<tr>
<td>No. regressors</td>
<td>9</td>
<td>5</td>
<td>14</td>
<td>7</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>No. dummies</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>0.657</td>
<td>0.702</td>
<td>0.718</td>
<td>0.798</td>
<td>0.671</td>
<td>0.813</td>
</tr>
<tr>
<td>No. regressors</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>No. dummies</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Both</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\sigma} )</td>
<td>0.553</td>
<td>0.753</td>
<td>0.712</td>
<td>0.819</td>
<td>0.767</td>
<td>0.788</td>
</tr>
<tr>
<td>No. regressors</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>No. dummies</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

\( \hat{\sigma} \) = equation standard error, No. regressors and No. dummies are number of regressors and, as a subset, number of dummies retained: Cons and Super are conservative and super-conservative strategies.
RMSFE averaged across horizon, IIS or not, and $\alpha$ for GDP and GDP growth.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factors</th>
<th>Both</th>
<th>RW</th>
<th>AR(D)</th>
<th>AR(I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \hat{y}_{T+k}$</td>
<td>0.666</td>
<td>0.600</td>
<td>0.650</td>
<td>0.666</td>
<td>0.485</td>
</tr>
<tr>
<td>$\hat{y}_{T+k}$</td>
<td>1.931</td>
<td>1.971</td>
<td>1.925</td>
<td>3.509</td>
<td>1.324</td>
</tr>
</tbody>
</table>

Castle, Clements & Hendry Forecasting by factors, by variables, by both, or neither? – p.58/76
Factor model best at short horizons but worst at long. Variables model improves in forecast accuracy at 8-step. Some benefit to IIS but few indicators retained at tight $\alpha (<6)$, roughly the same for factors and variables models. In levels, factor forecasts worsen as horizon increases. IIS yields greater improvements as retained dummies translate to level shifts in 1984.
Distributions of forecast errors

Factors

Variables

Both

Variables

Factors

Both

Castle, Clements & Hendry
Forecasting by factors, by variables, by both, or neither? – p.60/76
### In-sample results for inflation models with IIS

<table>
<thead>
<tr>
<th>Variables</th>
<th>1-step</th>
<th></th>
<th>4-step</th>
<th></th>
<th>8-step</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cons</td>
<td>Super</td>
<td>Cons</td>
<td>Super</td>
<td>Cons</td>
<td>Super</td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.240</td>
<td>0.283</td>
<td>0.293</td>
<td>0.290</td>
<td>0.227</td>
<td>0.296</td>
</tr>
<tr>
<td>No. regressors</td>
<td>20</td>
<td>12</td>
<td>17</td>
<td>13</td>
<td>21</td>
<td>12</td>
</tr>
<tr>
<td>No. dummies</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.263</td>
<td>0.336</td>
<td>0.299</td>
<td>0.318</td>
<td>0.311</td>
<td>0.327</td>
</tr>
<tr>
<td>No. regressors</td>
<td>15</td>
<td>7</td>
<td>20</td>
<td>14</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>No. dummies</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Both</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>0.266</td>
<td>0.284</td>
<td>0.302</td>
<td>0.356</td>
<td>0.284</td>
<td>0.334</td>
</tr>
<tr>
<td>No. regressors</td>
<td>14</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>No. dummies</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

$\hat{\sigma}$ = equation standard error, No. regressors and No. dummies are number of regressors and, as a subset, number of dummies retained: Cons and Super are conservative and super-conservative strategies.
RMSFE for inflation and its change

RMSFE averaged across horizon, IIS or not, and $\alpha$ for inflation and the change in inflation.

| Variables Factors Both RW AR(D) AR(I) |
|-----------------|-------|------|-------|-------|-------|
| $\Delta \hat{y}_{T+k}$ | 0.535 | 0.582 | 0.491 | 0.577 | 0.416 | 0.415 |
| $\hat{y}_{T+k}$ | 0.687 | 0.981 | 0.656 | 0.469 | 0.444 | 0.464 |

![Graphs showing RMSFE for different scenarios]
All models similar over forecast horizon but factor models worst on average as horizon increases. Small gains to IIS for variables or factor models. Tighter $\alpha$ improves forecasts, parsimonious models yield smaller RMSFE. Factor models most useful for shorter horizons. Levels highlight worse forecast accuracy of factors.
Distributions of forecast errors

Castle, Clements & Hendry Forecasting by factors, by variables, by both, or neither? – p.64/76
Does selection and estimation affect forecast accuracy? Compare four methods:

1. model is fixed after selection in-sample over \( t = 1, \ldots, T \);

2. model is selected over \( t = 1, \ldots, T \), parameters and eigenvalues recursively re-estimated over each forecast horizon, \( T = 1, \ldots, T + h \);

3. model is both selected and re-estimated recursively over each forecast horizon, \( t = 1, \ldots, T + h \), with eigenvalues re-calculated, so specification can change with each new observation; and

4. partial recursive selection: previously retained variables are not dropped even if insignificant and selection is only over additional variables or factors.
Castle, Clements & Hendry Forecasting by factors, by variables, by both, or neither? – p.66/76
(1) Setting the scene
(2) Relating variables $\{z_t\}$ to factors $\{f_t\}$
(3) Comparing variable-based and factor-based models
(4) Taxonomy of forecast errors in factor models
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusions
Can jointly include principal components and variables. Automatic selection allows $N > T$, perfect collinearities and multiple breaks. *Autometrics* can select models from both PCs and variables. Variables models and factor models are EqCMs, so non-robust to location shifts. Impulse-indicator saturation can account for shifts in-sample, and could be used for intercept corrections at forecast origin. Find some advantage to using IIS even for forecasting in differences: levels forecasts highlight the level shifts. Empirical results for GDP and inflation are mixed: factor models seem more useful for 1-step forecasts.
For multi-step forecasting, *Autometrics* selection over variables and PCs better than factor forecasts: benefits to selecting weights from correlations with $y_{t+h}$.

Little evidence that recursive estimation or recursive estimation and selection is better.

Gains to using IIS at tight $\alpha$ controlling null retention rate.

Selection jointly on factors and variables avoids imposing specific form, circumvents arbitrary choice of factors.

Whether data better described by factors or variables depends on phenomena, but can be determined by data using model selection techniques with low cost from retaining irrelevant effects.
References


(1) Setting the scene
(2) Relating variables, $\{z_t\}$, to factors $\{f_t\}$
(3) Comparing variable-based and factor-based models
(4) Taxonomy of forecast errors in factor model
(5) Forecast failure in equilibrium-correction models
(6) Model selection with both factors and variables
(7) Forecasting US GDP and inflation

Conclusion
To tackle multiple breaks & data contamination (outliers), add $T$ impulse indicators to candidates for $T$ observations.

Consider $x_i \sim \text{IID } [\mu, \sigma^2_\epsilon]$ for $i = 1, \ldots, T$

$\mu$ is parameter of interest

Uncertain of outliers, so add $T$ indicators $1_{\{t=t_i\}}$ to set of candidate regressors.

First, include half of indicators, record significant: just ‘dummying out’ $T/2$ observations for estimating $\mu$

Then omit, include other half, record again. Combine sub-sample indicators, & select significant.

$\alpha T$ indicators selected on average at significance level $\alpha$

Feasible ‘split-sample’ impulse-indicator saturation (IIS) algorithm
Johansen and Nielsen (2009) extend IIS to both stationary and unit-root autoregressions. When distribution is symmetric, adding $T$ impulse-indicators to a regression with $n$ variables, coefficient $\beta$ (not selected) and second moment $\Sigma$:

$$T^{1/2}(\tilde{\beta} - \beta) \xrightarrow{D} N_n \left[0, \sigma^2 \Sigma^{-1} \Omega_{\beta}\right]$$

Efficiency of IIS estimator $\tilde{\beta}$ with respect to OLS $\hat{\beta}$ measured by $\Omega_{\beta}$ depends on $c_\alpha$ and distribution. Must lose efficiency under null: but small loss $\alpha T$—only 1% at $\alpha = 1/T$ if $T = 100$, despite $T$ extra candidates. Potential for major gain under alternatives of breaks and/or data contamination: variant of robust estimation but can be done jointly with all other selections.
Size of the break is 10 standard errors at 0.75T

There are no outliers in this mis-specified model as all residuals $\in [-2, 2]$ SDs: outliers $\neq$ structural breaks

step-wise regression has zero power

Let’s see what Autometrics reports
Null ‘split-sample’ search in IIS

Dummies included initially

Dummies retained

Selected model: actual and fitted

Block 1

Block 2

Final

Castle, Clements & Hendry Forecasting by factors, by variables, by both, or neither? – p.75/76
‘Split-sample’ search in IIS

Dummies included initially

Dummies retained

Selected model: actual and fitted

Block 1

Block 2

Final

Castle, Clements & Hendry

Forecasting by factors, by variables, by both, or neither? – p.76/76