Do core inflation measures help forecast inflation?  
Out-of-sample evidence from French data

Hervé Le Bihan*, Franck Sédillot¹

Banque de France, 41-1391 DGEI-DEER, 31 rue Croix des Petits Champs, 75001 Paris, France

Received 27 August 1999; accepted 25 April 2000

Abstract

This paper compares the ability of four indicators of underlying or ‘core’ inflation to forecast inflation in the French case. Though most indicators Granger-cause inflation, results from out of sample tests of forecast accuracy are less compelling. The results nevertheless seem to give some empirical support to trimmed mean indicators. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Core inflation; Forecasting inflation

JEL classification: C4; E5

1. Introduction

A desirable property of an indicator of core inflation is that it should contain some information about future inflation. In a recent paper Freeman (1998) tests the information contents of core inflation on US data using Granger causality tests.

The present paper proposes an assessment of the forecasting accuracy of core inflation indicators using formal out of sample tests of predictive accuracy in addition to Granger causality tests. We also consider a wider set of indicators of core inflation than Freeman (1998). Our dataset is the monthly French CPI over the 1975–1998 period.

The remaining of the paper is structured as follows. The alternative indicators are briefly introduced in Section 2. Section 3 presents our methodology. Section 4 discusses the results of Granger causality tests. The forecasting performance of the various indicators with respect to future inflation is assessed using out-of-sample simulations in Section 5. Section 6 concludes.

*Corresponding author. Tel.: +33-142-924-987; fax: +33-142-926-292.
E-mail address: herve.lebihan@banque-france.fr (H. Le Bihan).
¹Banque de France. The views expressed in this paper are not necessarily those of the Banque de France.
2. Indicators of core inflation

Four broad measures of core inflation are included in our analysis:
(1) the traditional excluding food and energy inflation measure (EXC);
(2) limited influence estimators, in the spirit of Bryan and Pike (1991), Bryan and Cecchetti (1994) and Cecchetti (1996). Note that we use the 60th percentile (P60) rather than the median inflation in order to have average inflation corresponding to average core inflation. We also compute a trimmed mean estimator (TRIM) based on an asymmetrical trim (10% is removed from the left tail and 5% from the right tail);
(3) The Structural VAR measure (SVAR) computed according to the methodology of Quah and Vahey (1995);
(4) Last, the Dynamic Factor Index (DFI) measure, proposed by Bryan and Cecchetti (1993), which is based on an unobserved component model.

The construction of those indicators is presented and discussed in Le Bihan and Sédillot (1999).

3. Comparing forecast accuracy: methodology

To assess the relative predictive accuracy of the indicators we adopt a simple framework: bivariate VAR models including actual inflation and each underlying inflation indicator.

Two types of tests are presented. First we run Granger causality tests, which amount to test whether each of the indicators improves the in sample one-step ahead forecast of inflation. Secondly we perform the Diebold-Mariano (DM) (1995) test of comparison of forecast accuracy at various horizons. We use an autoregressive (AR) model of inflation as a benchmark. Let \( d \) be the difference of the Mean Squared Forecast Error (MSFE) between the AR model and the inflation-core inflation VAR model, and \( \hat{f}_d(0) \) an estimator of its spectral density at frequency zero. The DM statistics is:

\[
S_1 = \sqrt{\frac{d}{2\pi \hat{f}_d(0)}}
\]

Under the null hypothesis that the two models have the same forecast performance (i.e. \( E(d) = 0 \)) the test statistic \( S_1 \) is asymptotically normally distributed with unit variance. We use:

\[
2\pi \hat{f}_d(0) = \hat{f}_{0,T} + \sum_{j=1}^{m} \left[ 1 - \left( \frac{j}{m+1} \right) \right] \left( \hat{f}_{j,T} + \hat{f}_{j,T}' \right)
\]

with \( \hat{f}_{0,T} = (1/T) \sum_{t=1}^{T} (d_t - \bar{d})(d_{t-1} - \bar{d}) \). In the application \( m \) is chosen equal to \( 4(T/100)^{2/9} \). That the DM test allows for autocorrelation in the forecast error is particularly important. Indeed two

\footnote{It should be noted nevertheless that this exercise is not genuinely out-of-sample, since SVAR and DFI indicators used in the experiment are computed in the first stage using the whole sample information.}

\footnote{Using an asymptotic result is here relevant given the size of forecast errors series.}
successive observations of the cumulated forecast error should be mechanically correlated, since they are both moving average of monthly forecast errors.

The VAR model specification should take into account any non-stationarity in the data. Two competing hypotheses might be considered relevant, which appear difficult to distinguish from the data. The first is that inflation contains a unit root. The second is that it follows a stationary process, with presumably a break in the mid-80s, corresponding to the implementation of a disinflation policy in France. In the first case a sensible a priori restriction is that there is one cointegrating relationship between inflation and core inflation with a (known) unit coefficient. In such a case, the usual F test for Granger-type causality remains valid, in spite of the presence of a unit root, when performed on the unrestricted VAR (Sims et al., 1990). Causality tests can therefore in our case be run without deciding between a stationarity or cointegration assumption. To perform efficient forecasts it is nevertheless preferable to do so. Given the difficulty of distinguishing between the two hypotheses, we rely on the strategy suggested by Hamilton (1994, p. 652): run the experiment under both hypotheses, and assess whether similar answers emerge. We therefore conduct out of sample tests under the two following specifications: a VAR in level for the period 1985:1 to 1998:12 and a VECM for a longer time period (1975:1 to 1998:12), the former corresponding to stationarity-with-break hypothesis for the inflation process.

4. Granger-causality tests

For each of the core inflation indicators, we estimated a VAR in level for the period 1975:1 to 1998:12. The lag length is 6 (which is the length selected by both the Akaike and Schwartz information criteria). The VARs include seasonal dummies. Results are reported in Table 1. According to these in-sample tests, all indicators of underlying inflation convey some information about future inflation, except the SVAR measure. The results are unchanged when estimation is carried over a shorter time period. The poor performance of the structural VAR indicator is not surprising since this indicator intends to capture long run inflation (Quah and Vahey, 1995).

Table 1
Granger causality tests

<table>
<thead>
<tr>
<th></th>
<th>Exc.</th>
<th>Trim</th>
<th>DFI</th>
<th>P60</th>
<th>SVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975:1–98:12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F(6,256)$</td>
<td>4.8</td>
<td>4.8</td>
<td>3.6</td>
<td>4.6</td>
<td>1.1</td>
</tr>
<tr>
<td>$P$-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>1985:1–98:12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F(6,144)$</td>
<td>5.2</td>
<td>4.3</td>
<td>5.2</td>
<td>3.8</td>
<td>0.4</td>
</tr>
<tr>
<td>$P$-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.89</td>
</tr>
</tbody>
</table>

To hold, this result requires some specific assumptions on the number of cointegration relationships and variables involved in those relations (Toda and Phillips, 1993) which do hold in the present simple bivariate context.
5. Out-of-sample forecasting accuracy

Our approach is to recursively estimate a VAR model on samples beginning in 1985:01 and recursively ending from 1989:12 to 1995:12. For each estimate, a forecast is conducted simulating the VAR 36 months ahead. Thus, for every indicator of core inflation, we have 72 forecast sets, each containing 36 forecast horizons. For a given forecast horizon, the alternative indicators can be compared computing the MSFEs. As a benchmark, we also recursively estimate an AR model of inflation.

We report the results for selected horizons ($k=1$, 3, 12, 24, 36 months). For each horizon, our interest is in the forecast error in the overall inflation during the $k$ months ahead, i.e. the sum of the individual forecast errors for the horizons 1 to $k$. Table 2 presents the MSFEs based on cumulative forecast errors, relative to the MSFE of the AR model.

The poor predictive performance of the SVAR indicator is confirmed. According to the DM test, the SVAR indicator significantly outperforms the AR for the 36 months horizon only. This was to be expected since this indicator allows quite wide short-run deviations between core and actual inflation. Nor do inflation excluding food and energy or the DFI seem to convey valuable forward-looking information. On the other hand, for nearly every horizon, the 60th percentile and the trimmed mean perform better than other indicators. But they are significantly better than an AR model only for the horizons of 12 months and beyond.

We applied the same computation to the case of the cointegrated VAR. The VECM is then estimated without constant, with the number of cointegration relations set to one. Estimation period starts in 1975:1 rather than 1985:1 since we assumed the cointegration model is associated with the hypothesis of no break in the inflation process. The recursive sample periods used for forecasting are the same as above. In every case, the hypothesis of one cointegration relation is supported by the Johansen Trace test, and the estimated cointegration parameter is not significantly different from the expected value of 1.

Table 3 presents the results. It appears that MSFEs for the VECM are in general greater than for the VAR in level specification, except for the 3 years horizon. Thus on the sample period, forecasting inflation in level seems to bring better results. Furthermore no indicator outperforms the level autoregression. The AR model even significantly outperforms the SVAR for half of the horizons. These results might be strongly dependent on the particular estimation period, and on the sample period considered for assessing the forecast (1990:12 to 1998:12). When the VECM is estimated from

<table>
<thead>
<tr>
<th>$k$ months</th>
<th>Exc.</th>
<th>DFI</th>
<th>Trim</th>
<th>P60</th>
<th>SVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.02</td>
<td>0.96</td>
<td>0.97</td>
<td>0.93</td>
<td>1.05</td>
</tr>
<tr>
<td>3</td>
<td>1.11**</td>
<td>1.11**</td>
<td>0.97</td>
<td>0.91</td>
<td>1.04</td>
</tr>
<tr>
<td>12</td>
<td>1.10</td>
<td>1.12**</td>
<td>0.72**</td>
<td>0.73**</td>
<td>0.88</td>
</tr>
<tr>
<td>24</td>
<td>1.03</td>
<td>1.02</td>
<td>0.64**</td>
<td>0.68**</td>
<td>0.89</td>
</tr>
<tr>
<td>36</td>
<td>0.98</td>
<td>0.98</td>
<td>0.63**</td>
<td>0.68**</td>
<td>0.88*</td>
</tr>
</tbody>
</table>

*Significantly different from 1 at the 10% level for the DM test. MSFEs computed from cumulated forecast errors.

**Significantly different from 1 at the 5% level. MSFEs computed from cumulated forecast errors.
Table 3
MSFE relative to AR model’s MSFE VECM

<table>
<thead>
<tr>
<th>k months</th>
<th>Exc.</th>
<th>DFI</th>
<th>Trim</th>
<th>P60</th>
<th>SVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.19*</td>
<td>1.16</td>
<td>1.19*</td>
<td>1.16</td>
<td>1.39**</td>
</tr>
<tr>
<td>3</td>
<td>1.32</td>
<td>1.17</td>
<td>1.07</td>
<td>1.14</td>
<td>1.47**</td>
</tr>
<tr>
<td>12</td>
<td>1.38</td>
<td>1.22</td>
<td>1.25</td>
<td>1.23</td>
<td>1.57*</td>
</tr>
<tr>
<td>24</td>
<td>1.08</td>
<td>1.02</td>
<td>1.26</td>
<td>1.06</td>
<td>1.31</td>
</tr>
<tr>
<td>36</td>
<td>1.00</td>
<td>0.87</td>
<td>1.23</td>
<td>0.94</td>
<td>1.05</td>
</tr>
</tbody>
</table>

* Significantly different from 1 at the 10% level for the DM test. MSFEs computed from cumulated forecast errors.
** Significantly different from 1 at the 5% level. MSFEs computed from cumulated forecast errors.

1985:1 we get closer results to those in Table 2. We nevertheless regard modelling inflation as both I(1) and subject to a break in the 1980s as unattractive.

6. Conclusion

This paper has compared in the case of France the predictive ability of four indicators of underlying inflation: inflation excluding food and energy, trimmed mean, the structural VAR approach and the Dynamic Factor Index.

Though most indicators Granger-cause inflation, results from agnostic out of sample tests of forecast accuracy are less compelling. Few of the tested specifications significantly outperform the AR model. The results nevertheless seem to provide some empirical justification in favour of trimmed mean indicators, more particularly at the 12 month and beyond horizons. Forecasting inflation in level rather than with an inflation-core inflation VECM seems to bring better results over the 90s.

It should of course be reminded that designing a statistically efficient inflation forecast would require considering many alternative relevant variables (as in Stock and Watson, 1999, for instance).

References