The French order index backcasting through the use of cointegration theory

Laïla Bentoudja[1], Cynthia Gaborieau-Faivre[2]
[1]SSP, e-mail: laila.bentoudja@agriculture.gouv.fr
[2]Insee, e-mail: cynthia.faivre@insee.fr [1]

Abstract

A strong need for economic indicators in European countries that would help especially the ECB to manage its task led in 2001 to the STS (Short Term Statistics) Regulation. This regulation was made compulsory for European countries to compute monthly indicators on three main fields: industrial production, turnovers and orders.
France had then to cope with many problems because very few elements were available for orders. Worse, France was due to compute an index based in 2000, whereas none but a few activities orders had been collected then.
Of course the order index that was first computed in 2003 lacked robustness, mainly stemming from estimations. Worse, it was soon noticed that it showed undesirable volatility, precluding consistent seasonal adjustment. For instance in november 2005, the orders registered in travel equipment which used to be very smooth suddenly increased in an unbelievable fashion.
Last, ships appeared to have a much bigger impact on the index than trains did for equivalent amounts.
All of this led to the renewal of the index that was achieved in July 2007. It meant a threefold operation. First, giving up the weights of the index so as to avoid distorsions between activities. Second, seeking for any huge order such as that of november 2005 that had occured in the past and restoring it. And third, making abundant use of the cointegration theory in order to backcast the observed series.
The backcasting operation was done at the most elementary level. It involved invoices series as well as customs ones to play the role of safeguards in order to “drive” the order series through the past. Any divergence from the long-run relation between orders on the one hand and invoices or exports on the other hand was mechanically erased in a dynamic way.
The renewal of the index resulted in far more homogeneous series, allowing to maintain good quality in seasonal adjustment so as not to blur any more the message conveyed.

Keywords: [order index backcasting cointegration STS]

1. The context

After the release of the Short Term Statistics regulation in march 2001, a French new order index was bound to be built within a very short span of time. It had to cover the

[1] We thank Philippe Brion, Emmanuel Gros, Johara Khélif and Jean-François Loué for their help and advice

industry and construction fields rather extensively and was due to measure whole orders as well as orders intended for France, orders to be shipped to foreign areas, and Euro zone as well as non Euro zone orders also later on. That meant gathering or estimating an important amount of data. Yet the latter possibility (estimating) was prevailing in fact as very few series had been collected for a sufficiently long period (say since the year 2000 which had to be the base of the index).
The estimations were conducted through a mere use of evolutions observed on turnovers and therefore lacked robustness.
In 2007, the question of backcasting once more the order index, thanks to econometric methods this time, arose, since the data collected then covered a minimum of three years. At that time a considerable work was undertaken.

2. The work undertaken

The new series derived may show great differences with the previous ones. Indeed the aim of the French order index renewal was threefold.

First, the index weights _ the index is a Laspeyres one _ turned out to be inadequate, sometimes overweighting an activity and another times underweighting it. That led to giving them up, and to just weighting the orders by the average orders of the year 2000 themselves, even though the index coverage of the different activities was partial and depended on the activity considered.
Second, there soon appeared to be a scale problem with regard to the transportation orders, as a huge order was collected in November 2005 which could not be measured against previous orders. That meant that the device for collecting orders had failed in the past to measure such huge orders which had undoubtedly already occured. Therefore the press was extensively examined, seeking for their announcement.

Last but not least, a considerable amount of series was backcasted once more following an econometric method, say the cointegration theory, which led to more homogeneous time series in terms of volatility and offered better seasonal adjustment properties. The underlying idea behind was to make abundant use of already known series sharing common trends with orders, as far as it was possible. That was indeed the key point in the modelling, and a graphic inspection was done at first in order to give an idea of the potential success of the whole operation. One cannot make sure by this way that the modelization will be fruitful, yet in some cases it can appear useless, for instance if the series do not exhibit any similarity. Indeed one series should never drift away from the other one in a stochastic way and should always resume its convergence towards the other one by means of a “correction” of the corresponding error. Sometimes the introduction of a constant and a trend can be induced by the fact that part of the field covered by the index does not work properly according to the orders recorded. Moreover some invoices may be cancelled after they have been received.

3. How the backcasting operation was carried out

Both turnovers and invoices were presumably good candidates for backcasting orders, yet the latter ones were preferred as they were already present in the good nomenclature version (Naf Rev 1). The exports transmitted by the Customs were used as well. The whole orders on the one hand and the orders designed for exports on the other hand were backcasted, thus leading to the French orders by difference. If ever French orders happened to be negative, then they were replaced by zero. Fortunately that happened very seldom. As for Euro and none Euro zone orders, no long time series was available, so that their yearly evolution was defined as the product of the evolution of their parts given within the national accounts and of the export orders evolutions.

<table>
<tr><td rowspan="3"><u>Whole orders</u>
New backcasting operation</td><td>French orders

Difference between whole and foreign orders</td><td></td></tr>
<tr><td rowspan="2"><u>Foreign orders</u>
New backcasting operation</td><td>Orders for Euro zone

The part of European orders among foreign ones in the national accounts is used</td></tr>
<tr><td>Orders for non Euro zone

The part of non European orders among foreign ones in the national accounts is used</td></tr>
</table>

Table 1: the connections between the different backcasting operations

It was necessary to isolate the seasonal part of all the modelled series. To do so, dummy variables may be incorporated in the models in order to capture the seasonal effect, or seasonally adjusted series may be modelled before the raw data are calculated by way of combining the backcasted data with the average seasonal component. The latter solution was favoured.

4. Models used

The dynamics of the most general model is as follows:

$$\Delta Y_t = u_1 \Delta Y_{t-1} + \mu + \gamma * t - \Pi Y_{t-1} + \varepsilon_t$$

where Y stands for the vector of orders and invoices (respectively exports) taken in logarithm so as to stabilize them. Everything then depends on the rank of the matrix Π.

When Y is I(1) which means that its differential is stationary, the formulation above is quite satisfactory. If ever it is not, but the differential of its differential is stationary, then the right formulation is:

$$\Delta^2 Y_t = u_1 \Delta Y_{t-1} + \mu + \gamma * t - \Pi Y_{t-1} + \varepsilon_t$$

which involves both the acceleration rates and the growth rates also and therefore covers periods in which changes occur very fast as well as smoother periods.
Having noticed that the two models are equivalent, we will now focus on the first one.

If the rank of Π is null then it is obvious that Π is null also, and so a mere model in differences is postulated:

$$\begin{pmatrix} \Delta orders_t \\ \Delta invoices_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} * t + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} * \begin{pmatrix} \Delta orders_{t-1} \\ \Delta invoices_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

On the contrary if Π is of full rank, then it is obvious that Y must be also of full rank, thus a Var (Vectorial autoregressive regression) model on the levels is convenient:

$$\begin{pmatrix} orders_t \\ invoices_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} * t + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} * \begin{pmatrix} orders_{t-1} \\ invoices_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

We will focus here on the specific case when Π is not of full rank, yet not being null.

Y is non stationary, which means that it may incorporate a stochastic trend as well as a deterministic one. Yet ΠY_{t-1} is stationary, which yields that any stochastic trend in Y must lie in the orthogonal part of the cointegration space. As for the constant and the trend that appear in the general equation, they can also successfully be represented as an element lying in the cointegration space and an element lying in its orthogonal space. Thus the most general equation becomes:

$$\Delta Y_t = u_1 \Delta Y_{t-1} + \mu + \gamma^1 * t - \Pi \begin{bmatrix} orders_{t-1} \\ invoices_{t-1} \\ 1 \\ t-1 \end{bmatrix} + \varepsilon_t$$

which can be developed as:

$$\begin{pmatrix} \Delta orders_t \\ \Delta invoices_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \gamma^1{}_1 \\ \gamma^1{}_2 \end{pmatrix} * t + \Gamma * \begin{pmatrix} \Delta orders_{t-1} \\ \Delta invoices_{t-1} \end{pmatrix} + \Pi \begin{bmatrix} orders_{t-1} \\ invoices_{t-1} \\ 1 \\ t-1 \end{bmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

(Equation A)

which is estimated by a likelihood maximization method. Twelve parameters have to be estimated then. Let us explain where the model stems from in an intuitive way.
Cointegration relations often also called "long-run equilibria" act as an attractor, so that any departure therefrom will always be erased in the future. They may incorporate a deterministic trend if the deterministic trends present in the data do not naturally cancel in the linear combination, and also a constant. The vector of growth rates in the left-hand side and the other one in the right-hand side explain how the new balance is achieved in a dynamic way thanks to the attraction to the cointegration relation.

From the last formulation of the model (A) five cases may be derived depending on the trends and constants that are incorporated [2]:

- Case 1: No restriction on the deterministic terms. Then the model is consistent with a linear trend in the differenced series and so may apply to quadratic series. Such a model must be used with caution, since it may improve the fit but can produce unexpected results when used for forecasting.
- Case 2: γ^1 =0, therefore the trend lies in the cointegration space. So the model is convenient for series with a trend, but not for quadratic ones.
- Case 3: there is no trend at all in the model (γ =0). This model is consistent with non-centered differentials of the series.
- Case 4: the model precludes any trend in the data but would be consistent with a constant should it lie in the cointegration space (γ =0 and $\mu = 0$).
- Case 5: the model precludes any deterministic term. This can hardly be consistent with real data.

Here we will at first model the series as in Equation (A). The Johansen cointegration theory can help to derive the most convenient model.
First, a Johansen cointegration test is performed.

4.1 Tests performed

Two tests for determining the rank of the Pi matrix are available, which are:

the λ -trace test and the λ -max test.
For the λ -trace test, the null hypothesis is as follows:

(H0): $r_0 \leq r$ (where r_0 stands for the rank of the Π matrix) whereas the alternative is:

(H1): $r_0 > r$

[2] This presentation comes from Henry D. F., Juselius K. (2000).

As for the λ-max test, the null hypothesis is:

(H0): $r_0 = r$ whereas the alternative is:

(H1): $r_0 = r + 1$[3]

The critical values of these tests depend both on the number of variables involved and on the deterministic structure added (which has been summarized within the five cases depicted above).

4.2 Three models proposed

According to the value of the matrix Π we have:

A - When the rank of the matrix Π is 0, the general equation becomes:

$$\begin{bmatrix} DLorders(k) \\ DLinvoices(k) \end{bmatrix} = \begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & c(1,4) \\ c(2,1) & c(2,2) & c(2,3) & c(2,4) \end{bmatrix} \begin{bmatrix} DLorders(k-1) \\ DLinvoices(k-1) \\ time(k-1) \\ 1 \end{bmatrix}$$

Some coefficients may be taken off according to the option selected.

Here we have according to the deterministic structure chosen (case 1, ...):

	C(1,1)	C(1,2)	C(1,3)	C(1,4)
Case 1			=0	=0
Case 2			=0	
Case 3			=0	
Case 4				
Case 5				

B - When the Pi matrix has a rank equal to 1, the general equation becomes:

$$\begin{bmatrix} DLorders(k) \\ DLinvoices(k) \end{bmatrix} = \begin{bmatrix} a(1,1) \\ a(1,2) \end{bmatrix} \begin{bmatrix} b(1,1) & b(1,2) & b(1,3) & b(1,4) \end{bmatrix} \begin{bmatrix} Lorders(k-1) \\ Linvoices(k-1) \\ time(k-1) \\ 1 \end{bmatrix} +$$

$$\begin{bmatrix} C(1,1) & C(1,2) & C(1,3) & C(1,4) \\ C(2,1) & C(2,2) & C(2,3) & C(2,4) \end{bmatrix} \begin{bmatrix} DLorders(k-1) \\ DLinvoices(k-1) \\ 1 \\ time(k-1) \end{bmatrix}$$

[3] The conclusions of these tests may not coincide, which led to a preselection of a larger set of models following the AIC and Schwartz information criteria.

In that case, there is a cointegrating relation between orders and invoices: an error correcting model is being estimated. Some coefficients may be taken off depending on the option selected. Actually we have:

	A(1,1)	B(1,1)	B(1,2)	B(1,3)	B(1,4)	C(1,1)	C(1,2)	C(1,3)	C(1,4)
Option 1				=0	=0			=0	=0
Option 2				=0				=0	=0
Option 3				=0					=0
Option 4									=0
Option 5									

C - When the rank of the Pi matrix is 2, the general equation becomes:

$$\begin{bmatrix} Lorders(k) \\ Linvoices(k) \end{bmatrix} = \begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & c(1,4) \\ c(2,1) & c(2,2) & c(2,3) & c(2,4) \end{bmatrix} \begin{bmatrix} Lorders(k-1) \\ Linvoices(k-1) \\ time(k-1) \\ 1 \end{bmatrix}$$

Again we have another table:

	C(1,1)	C(1,2)	C(1,3)	C(1,4)
Option 1			=0	
Option 2			=0	
Option 3			=0	
Option 4				
Option 5				

In practice, several criteria have to be checked in order to ensure that the dynamics of the model is convergent when the model is an error correcting one, and to allow for its economic interpretation.
Indeed, the first estimations obtained may sound inadequate from an economic or a statistical point of view. Therefore they must be further checked:

- Firstly, the elasticity of orders to invoices should be exactly one.

We have actually accepted values ranging from 0 to 2.
As the chronological order has been reversed, the estimated model can be expressed as
P(L)(*invoices*) = Q(L)(*orders*), where P and Q stand for polynomials and L stands for the lag operator:
$L(Y_t) = Y_{t-1}$
When the model is written in growth evolution, we can write: P(L) = (1-L)P1(L) and Q(L) = (1-L)Q1(L) where P1 and Q1 are two polynomials.
The long term elasticity of orders to invoices can be simplified as Q(1)/(P(1) if the model is written in levels and Q1(1)/P1(1) if it is written in growth evolutions.

- Secondly, the dynamics of a model expressed in levels must be convergent:

(1-L)(*Iorders*)=αL[*Iorders*-P(1)/Q(1).(*invoices*)+b+c**TIME*]+
Q2(L)(1-L)(*invoices*)+P2(L)(1-L)(*Iorders*)

Then the convergence can occur if and only if α strictly falls between -2 and 0. It is monotonous only if α falls between -1 and 0.

- Thirdly, the mean lag of adjustment (MLA) must be acceptable.

That means that the average number of months (for monthly data) that elapse before a shock over orders is overcome in invoices may not exceed the average duration between an order and the end of the corresponding production, i.e. the length of the production cycle.

- MLA = P'(1)/P(1) - Q'(1)/Q(1), when the model is expressed in levels
- MLA = P1'(1)/P1(1) - Q1'(1)/Q1(1), when the model is expressed in growth evolutions.

Indeed if we have:

$$Y = P(L)X$$

Then we also have MLA = P'(1)/P(1). Let us explain that point:

$$MLA = \frac{weighted\_sum\_of\_the\_adjustment\_delays}{total\_adjustment}$$

which stems from its definition. So if we put:

$$Y_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p}$$

then we have:

$$MLA = \frac{0 * a_0 + 1 * a_1 + 2 * a_2 + 3 * a_3 + \cdots + p * a_p}{a_0 + a_1 + \cdots + a_p}$$

When formally identifying the derivative of P/Q to its "infinite" polynomial expression, the result can be derived.

The MLA is expressed in the same time unit than the rythm of the data record.

An abnormal figure for the mean lag of adjustment (in fact a figure bigger than 6) may be indicative of a bad specification.

Should the elasticity present a bad value, a mere autoregressive model would be adopted.

For other problems, they can be settled through an outlier detection and correction, and/or a time span reduction.

The strategy followed throughout the backcasting operation can be summed up as follows:

1 - Both invoices and the corresponding orders are seasonally adjusted
2 - The time direction is reversed
3 - The rank of cointegration is determined
4 - The model is being estimated conditionnally to the number of intependent cointegration relations
5 - The interpretability of the model is questionned
6 - The backcasting operation is under way
7 - The time direction is reversed once more
8 - The raw data are derived from the seasonally adjusted one.

5. The series covering orders in the steel industry as an example

By way of illustration, let’s take the example of the series which covers all orders received by the steel industry and due to be shipped to foreign countries. The span of time for which it has been collected runs from 2001.
Let’s first overlay both foreign orders and customs exports:

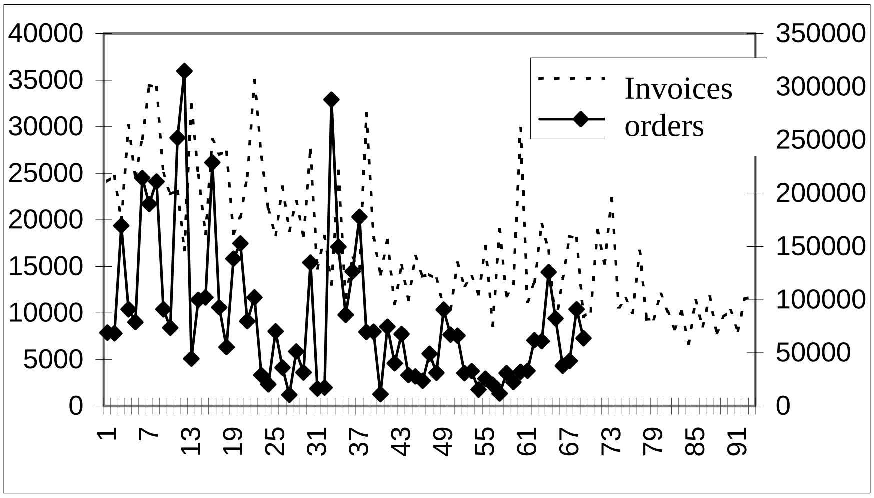

Figure 1**:** the graphical inspection let us hope the modelization will be fruitful

A strong parallelism show up, meaning that the modelling through cointegration should be fruitful.
Let’s then examine the different tests for the rank of cointegration.
We follow the parcimony principle whenever possible. Above all, the option number five which assumes that there’s a quadratic trend in the data should be avoided since it is very seldom consistent with economic explanations.
Here we pre-select the option 4 for rank 2 and the option 1 for rank 1 by minimizing the information criteria.
In this case it is the first model which is selected. Indeed it is the only one presenting a relatively good elasticity.
Therefore the equation is:

$$\begin{pmatrix} \Delta LICom_t \\ \Delta LF_t \end{pmatrix} = \begin{bmatrix} -0{,}1 & -0{,}1 \\ -0{,}3 & -0{,}5 \end{bmatrix} \bullet \begin{pmatrix} \Delta LICom_{t-1} \\ \Delta LF_{t-1} \end{pmatrix} + \begin{bmatrix} -0{,}6 \\ 0{,}2 \end{bmatrix} \bullet \begin{bmatrix} 1{,}0 & -1{,}13 & 0 & 0 \end{bmatrix} \bullet \begin{bmatrix} LICom_{t-1} \\ LF_{t-1} \\ 1 \\ t-1 \end{bmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

A very smooth backcasted series has often been the result of the backcasting operation which was conducted.
At last, after reversing the time direction, we had the following raw data which seemed sensible:

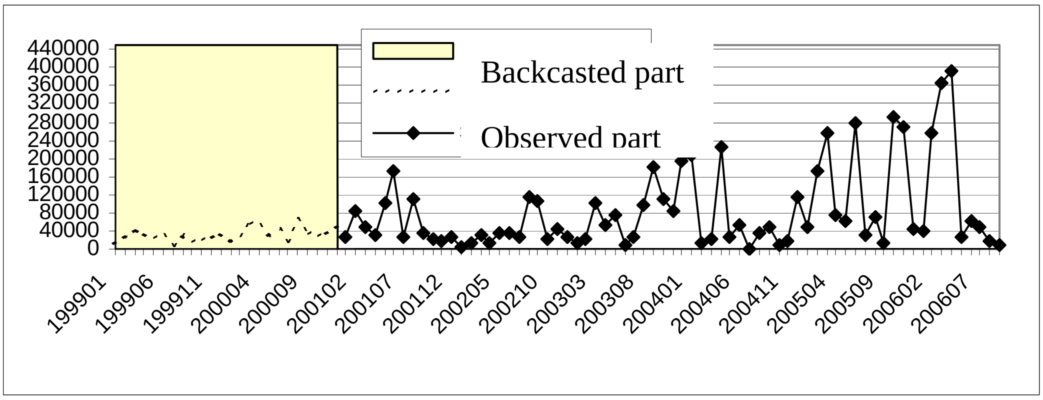

Figure 2: there is no level shift between the backcasted part and the observed one

Last, let's examine whether the use of external information has been intensive or not.

			Beginning of the					
			<2000	2001	2002	2003	≥2004	Total
Backcasting series	Econometric regression	Monthly invoices		32	15	3		*50*
		Quarterly invoices		6	11	-		*17*
		Monthly customs exports		52	35	3		*90*
		Autoregressive model		36	17	10		*63*
		IPI PPI			2			*2*
		Huge orders (press)				10		*10*
		Nace 2[4]		16	10		10	*36*
		Without point	192		2	4		*198*
		Total	*192*	*142*	*92*	*30*	*10*	*466*

Table 2: an intensive use of external information

A high degree of incorporation of information may be noticed.
Besides, the renewed index shows discrepancies with the last one because of its reduced volatility, especially for the transportation material activity. From now on it should show less erratic revisions on its seasonally adjusted series, since seasonality does not know any more severe changes between backcasted data and actually observed data. In the end the series are now much more consistent with the combined evolution of the French ipi and ppi than they were before.
Therefore the whole operation turns out to have been a way of seriously improving the quality of the index.

References

Bentoudja L., Gaborieau-Faivre C. (to be published in 2009) La rétropolation économétrique de l'indice de commandes français, *JMS*

Bernard A., Roy G. Cooper M. C. (2003) Etude des distorsions de niveau des tests de Johansen pour la cointégration, *Cahier de recherche 03-08 Département d'économie, Université de Sherbrooke*

Henry D. F., Juselius K. (2000) Explaining cointegration analysis: part II, *Discussion papers, Department of Economics, University of Copenhagen*

[4] When no relevant data is available for an elementary series, its backcasting operation should not contribute to the evolution of the aggregates: that is why the evolution of the aggregates of already backcasted data is applied.