"Infinite" data and few information for regional forecast: an applied approach from this paradox[*]

Paola M. Chiodini,
Department of Statistics, University of Milano-Bicocca,
Via Bicocca degli Arcimboldi, 8, Milan, Italy

Silvia Facchinetti,
Department of Statistical Science, Catholic University, Via Necchi, 9, Milano, Italy –
ESeC, Via Matteotti, 15/A, Assago (MI), Italy

Giancarlo Manzi,
Department of Economics, Business and Statistics, Università degli Studi di Milano,
Via Conservatorio, 7, Milan, Italy

Flavio Verrecchia[†],
ESeC no-profit Association, Via Matteotti, 15/A, Assago (MI), Italy,
e-mail: flavio.verrecchia@gmail.com

Abstract: Although the unstoppable evolution of Information Technology allows nowadays for the treatment of massive socio-economic data, reliable historical data for forecasting are not always widely available for applying statistical models from which significant parameters can be obtained. This becomes a paradox in regional forecasting where data can be widely available but often can be useless. The aim of this work is to present a method for improving forecasting precision with few information. After a brief description of the most popular techniques in this field, an application on regional forecasting with hierarchical conciliation is presented.

Keywords: Official statistics, Large-scale forecasting with few information, Joint use of forecasts with different time scaling, Regional data treatment.

1. Introduction and Aims

In sectorial and regional forecasting in the European Union (EU), although data are often widely available, it is customary to handle situations where data are not suitable for the following reasons: (i) administrative regional borders can have been modified; (ii) the classification of economic activities can have been revised; (iii) EU harmonization policies on survey domains can have been adopted.

In this paper, after a brief review of the most popular techniques used to forecast, an application on regional forecasting with hierarchical conciliation is presented in order to find new ways to handle regional forecasting in presence of few information. To this purpose, annual (regional data) and quarterly (country data) time series on the labour market, available from the Eurostat database, are used.

[†] Flavio Verrecchia is the correspondence author.

The joint analysis of multiple time series can be easily performed with statistical tools which, in the majority of cases, provide satisfactory solutions both in terms of model fit and in terms of forecast reliability. It is doubtless that this type of solution is desirable when researchers have to adjust classical estimates, in particular in cases when time series present some *breaks*, due for example to some macroeconomic factors which affect their stationarity.

In particular, if one wants to analyze Eurostat data at a regional level, the most used techniques are those of classical time series analysis (i.e. exponential smoothing) proposed in the 50s e 60s by Holt, Brown e Winters (see Holt, 1957; Brown (1959); Granger et al., 1986; Santamaria, 2000; Chatfield et al., 2001; Hyndman et al., 2008).

However, EU national survey agencies can use different survey methods or can introduce new survey methodologies in order to take into account the EU policies. Furthermore, although many improvements have been made, national survey agencies have not to date completed the harmonization process. These are among the reasons for which time series are not completely comparable and have to be truncated for forecasting. This problem arise also at a national level.

The methods we use to deal with the aforementioned problems on official statistics data are those of exponential smoothing expressed in terms of state space models (Hyndman et al. 2008). We apply these state space models together with reconciliation methods in order to conciliate regional forecast to national forecast, and, so doing, improved estimates are obtained. The reconciliation is performed at a national level where traditional statistical models provide more reliable estimates since data are usually based on a higher sampling rate.
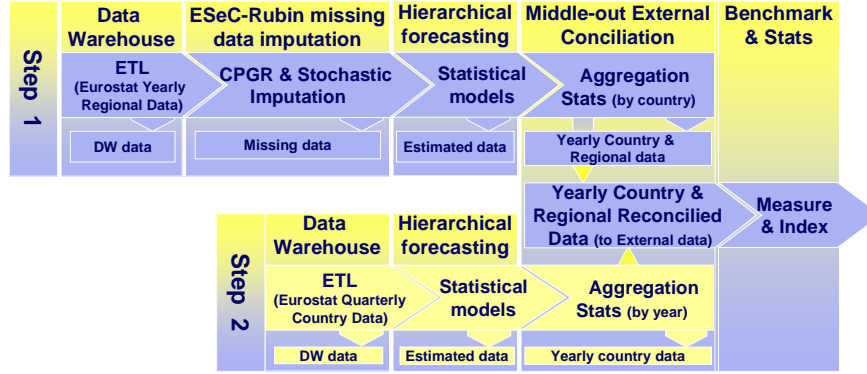

2. Methods

2.1. The External Middle-Out Hierarchical Forecasting

Multiple strategies are available for regional forecast aggregation. The bottom-up strategy is successful, for example, when it is applied to regional demography, where time series are longer, more complete and with low variability. However, if the quality of time series is not satisfactory, and/or macro-regional time series with shorter time scale are available, then one has to choose a different strategy - the External Middle-Out Hierarchical Forecasting (EMOHF) - which can be based on a joint use of multiple forecasts (for the same aggregate) and is developed in two steps. *Middle-Out* stands for bottom-up versus EU domain (Marcellino, 2004) and top-down to regional level.

This method can be applied above all to data not suitable for modelling. In these cases, even if the best model cannot be suitable from an inferential point of view (i.e. with no significant parameters), estimates can be improved. This issue can be generally solved through the use of auxiliary variables.

The EMOHF strategy consists of performing two separate forecasts, one at a regional and one at a national level, and then to obtain a *conciliation* to the external national data (Figure 1). This strategy overcomes the information gap through an initial estimate of the current regional data to which an adjustment is applied in order to get a conciliation with the external national data, which is therefore estimated starting from a different data set (i.e. a more recent quarterly data).

Figure 1: Double-phase National Middle-Out Hierarchical Forecasting strategy

Source: Verrecchia (2008).

2.2. State space models

All Exponential Smoothing methods automatically selected in our application can have a State space representation.

Let ℓ be the level term, b be the growth term, T_h denote the forecast term over the next h time periods and ϕ denote the damping parameter ($0 < \phi < 1$). ℓ and b can be combined given five future trend patterns:

- None (N): $T_h = \ell$
- Additive (A): $T_h = \ell + bh$
- Additive damped (A_d): $T_h = \ell + (\phi + \phi^2 + ... + \phi^h)b$
- Multiplicative (M): $T_h = \ell\, b^h$
- Multiplicative damped (M_d): $T_h = \ell\, b^{(\phi + \phi^2 + ... + \phi^h)}$

Having chosen the trend component, we have to match it with the seasonal component: none (N), additive (A) or multiplicative (M) (Table 1).

Table 1: Exponential Smoothing Methods

Trend component	Seasonal component		
	N	**A**	**M**
N	N,N	N,A	N,M
A	A,N	A,A	A,M
A_d	A_d,N	A_d,A	A_d,M
M	M,N	M,A	M,M
M_d	M_d,N	M_d,A	M_d,M

Source: Taylor (2003)

Then, considering the triplet E, T, S (Error, Trend, Seasonality), the automatically selected models in our application (Section 3) are:

1. Linear Exponential Smoothing and Double Exponential Smoothing (Brown) - ETS(A,N,N);
2. Linear Exponential Smoothing (Holt) - ETS(A,A,N);
3. Damped-Trend Linear Exponential Smoothing - ETS(A,A_d,N);
4. Additive Seasonal Smoothing (Winters) - ETS(A,A,A).

Let ℓ_t denote the series level at time t, b_t denote the slope at time t, s_t denote the seasonal component of the series at time t and m denote the number of seasons. Then is possible

to express the Exponential Smoothing equations (where α, β^*, γ, ϕ are constants; $\phi_h = \phi + \phi^2 + \ldots + \phi^h$ and $h^+_m = [(h-1) \bmod m] + 1$) (Table 2).

We assume that the errors are independent and identically distributed, following a Gaussian distribution with zero mean and variance equal to σ^2, $\varepsilon_t \sim NID(0,\sigma^2)$, the state space general equations are:

$$y_t = w(\boldsymbol{x}_{t-1}) + r(\boldsymbol{x}_{t-1})\, \varepsilon_t,$$
$$\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}) + g(\boldsymbol{x}_{t-1})\, \varepsilon_t,$$

where $\boldsymbol{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \ldots, s_{t-m+1})'$, $\mu_t = w(\boldsymbol{x}_{t-1})$ and with additive error $r(\boldsymbol{x}_{t-1}) = 1$.

Let $\mu_t = \hat{y}_t$ denote the one-step forecast for y_t and $\varepsilon_t = y_t - \mu_t$ denote the one–step forecast error at time t. Considering the triplet E, T, S, we can find the state space models for each exponential smoothing methods (to simplify the notation we use $\beta=\alpha\beta^*$) (Table 3).

Table 2: Exponential Smoothing Formulae

Methods	Equations	
N,N	$\ell_t = \alpha y_t + (1 - \alpha)\, \ell_{t-1}$	[1a]
	$\hat{y}_{t+h\|t} = \ell_t$	[1b]
A,N	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1}+ b_{t-1})$	[2a]
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\, b_{t-1}$	[2b]
	$\hat{y}_{t+h\|t} = \ell_t + h b_t$	[2c]
A_d,N	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1}+ \phi b_{t-1})$	[3a]
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\, \phi b_{t-1}$	[3b]
	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$	[3c]
A,A	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1}+ b_{t-1})$	[4a]
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\, b_{t-1}$	[4b]
	$s_t = \gamma(y_t - \ell_t - b_{t-1}) + (1 - \gamma)\, s_{t-m}$	[4c]
	$\hat{y}_{t+h\|t} = \ell_t + h b_t + s_{t-m+h^+_m}$	[4d]

Source: Hyndman R.J., Koehler A.B., Ord J.K, Snyder R.D. (2008).

Table 3: State space equations with additive error

Models	Equations	
ETS(A,N,N)	$\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	[5a]
	$\mu_t = \ell_{t-1}$	[5b]
ETS(A,A,N)	$\ell_t = \ell_{t-1}+ b_{t-1} + \alpha\varepsilon_t$	[6a]
	$b_t = b_{t-1} + \beta\varepsilon_t$	[6b]
	$\mu_t = \ell_{t-1} + b_{t-1}$	[6c]
ETS(A,A_d,N)	$\ell_t = \ell_{t-1}+ \phi b_{t-1} + \alpha\varepsilon_t$	[7a]
	$b_t = \phi b_{t-1} + \beta\varepsilon_t$	[7b]
	$\mu_t = \ell_{t-1} + \phi b_{t-1}$	[7c]
ETS(A,A,A)	$\ell_t = \ell_{t-1}+ b_{t-1} + \alpha\varepsilon_t$	[8a]
	$b_t = b_{t-1} + \beta\varepsilon_t$	[8b]
	$s_t = s_{t-m}+ \gamma\varepsilon_t$	[8c]
	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$	[8d]

Source: Hyndman R.J., Koehler A.B., Ord J.K, Snyder R.D. (2008). **Notes**: 1. ETS: Error, Trend, Seasonal component. 2. ETS(A,N,N): Linear Exponential Smoothing; Double Exponential Smoothing (Brown); ETS(A,A,N) Linear Exponential Smoothing (Holt); ETS(A,A_d,N): Damped-Trend Linear Exponential Smoothing; ETS(A,A,A): Additive Seasonal Smoothing (Winters).

3. Applications

In this section an application of the above models to the employment level in the Italian regions will be presented. Data are taken from the Eurostat database – Labour Force Survey section. In this database regional and sectorial aggregates are available with an annual periodicity. This fact constitutes a problem which can not be solved in terms of statistical models specification with significant parameters. This often happens also for models applied to national quarterly data. This is due to technological changes both in terms of innovation and in terms of harmonization. The proposed strategy is that of improving the estimates through the conciliation methodology, starting from the best models detected. In the following, the models will be specified by aggregating and conciliating the data. Finally, thanks to auxiliary information, a national model will be specified.

Automatic models specification

In our application we used SAS Forecast Server which can easily handle the hierarchical information and the automatic selection of models, on the basis of fit statistics (i.e. Mean Absolute Percentage Error - MAPE). It allows for the detection of regional and national models (see Tables 4 and 5).

It can be noted that regional automatically selected models result in some cases in a non-significant set of parameters. However, by comparing the Absolute Percentage Errors (APE) of the best automatically specified regional models with the APEs of the naïve predictors (average and not-centred moving average of five terms), it can be noted that in 2007, while for automatically specified regional models the APE does not exceed 4.5% (8.6% in 2008), for naïve predictors the APEs grow reaching more than 7% (9% in 2008) (Figure 2).

Table 4: Employment - persons aged 15-64 (thousands), regional models, forecasts, ex-post APEs, by Nuts2 (regions), 2007-08

| | Models | | Model parameters | | | | Weight / Damping | | Estimates | | | |
| | | | Level | | Trend | | | | 2007 | | 2008 | |
Nuts	Model ETS(…)	MAPE	Par. estim.	P-values	Par. estim.	P-value	Par. Estim.	P-value	$\hat{y}$ (000)	APE	$\hat{y}$ (000)	APE
ITC1	A,A,N	0.82%	0.132	0.323	0.001	0.997			1,834	0.2%	1,848	0.1%
ITC2	A,N,N	0.93%					0.971	0.000	56	0.5%	56	0.6%
ITC3	A,A,N	1.06%	0.001	0.998	0.001	1.000			622	2.0%	627	1.3%
ITC4	A,A,N	0.26%	0.276	0.049	0.001	0.975			4,262	0.7%	4,323	1.2%
ITD1	A,A,N	1.11%	0.999	0.013	0.001	0.996			224	0.1%	226	1.2%
ITD2	A,N,N	1.56%	0.999	0.006					216	1.9%	216	3.6%
ITD3	A,A,N	0.46%	0.205	0.158	0.001	0.992			2,090	0.2%	2,118	0.4%
ITD4	A,A,N	0.85%	0.129	0.342	0.001	0.998			512	0.3%	517	0.7%
ITD5	A,A,N	0.45%	0.193	0.177	0.001	0.993			1,891	1.1%	1,913	1.0%
ITE1	A,A,N	0.51%	0.188	0.204	0.001	0.994			1,524	0.6%	1,542	0.0%
ITE2	A,A,N	0.86%	0.047	0.730	0.001	1.000			349	3.1%	354	4.1%
ITE3	A,A,N	0.38%	0.267	0.166	0.001	0.986			642	0.4%	650	0.8%
ITE4	A,A,N	0.55%	0.999	0.020	0.001	0.996			2,126	2.5%	2,160	2.4%
ITF1	A,A,N	0.91%	0.216	0.221	0.001	0.993			501	1.2%	509	0.3%
ITF2	A,A,N	1.04%	0.001	0.993	0.001	1.000			108	2.5%	109	3.8%
ITF3	A,A$_d$,N	1.24%	0.209	0.685	0.001	0.999	0.999	0.000	1,771	4.1%	1,802	8.6%
ITF4	A,A,N	1.05%					0.999	0.000	1,280	0.8%	1,316	3.4%
ITF5	A,A,N	1.35%	0.048	0.723	0.001	1.000			196	1.7%	198	2.7%
ITF6	A,A$_d$,N	1.20%	0.192	0.665	0.001	0.999	0.999	0.000	624	4.5%	636	7.9%
ITG1	A,A,N	0.58%	0.179	0.182	0.001	0.994			1,500	1.9%	1,525	4.2%
ITG2	A,A,N	1.48%	0.179	0.162	0.001	0.993			615	1.6%	630	4.5%

Source: ESeC estimates on Eurostat data (Labour Force Survey) – forecast data at 08, ex-post at 07.

Notes: 1. Data for annual model assessment from 1999 to 2006 (2007-2008 forecasts). 2. $\hat{y}$: forecast. 3. APE (ex post): $|100 \ (y_t - \hat{y}_t)/y_t|$. 4. ETS(A,N,N): Linear Exponential Smoothing; Double Exponential Smoothing (Brown); ETS(A,A,N) Linear Exponential Smoothing (Holt); ETS(A,A$_d$,N): Damped-Trend Linear Exponential Smoothing.

Figure 2: Employment - persons aged 15-64, ex-post APEs, by Nuts2 (regions), '07-08
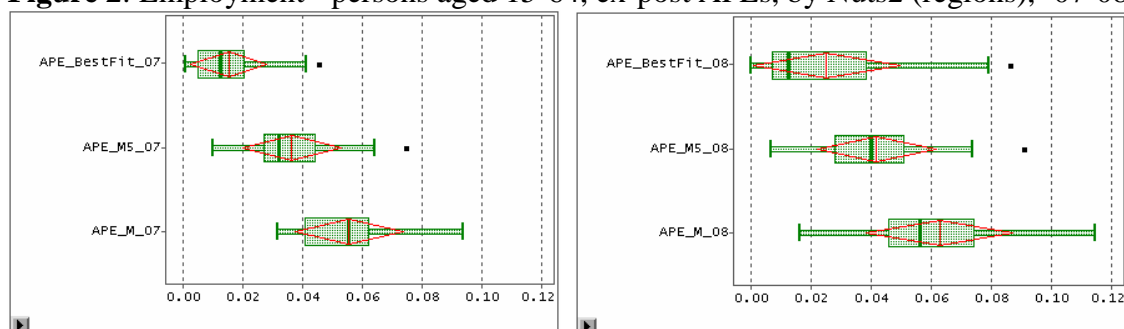
Source: ESeC estimates on Eurostat data (Labour Force Survey). **Notes**: 1. APE_BestFit: APE of estimates (Table 4); 2. APE_M: APE of mean; 3. APE_M5: APE of moving average of 5 terms.
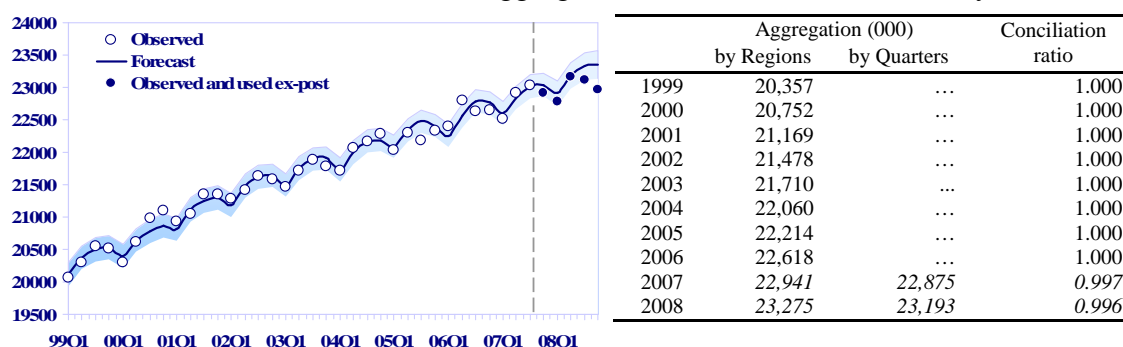
Aggregation and conciliation

As we have seen, regional automatically selected models result in some cases in a non-significant set of parameters. Nevertheless, it is possible to improve (or to make more robust) the regional estimates by conciliating them to national estimates. SAS Forecast Server allows for the aggregation and conciliation of the estimates. For example, the Italian employment level derived from the aggregation of regional estimates is overestimated (+0.3%) if compared with the forecast obtained as a quarterly aggregation of external national estimates (Table 5 and Figure 3).

Table 5: Employment - persons aged 15-64 (thousands), national model, ex-post APE, by Nuts0, 2007-08

Models			Model parameters						Estimates			
			Level		Trend		Seasonal		2007		2008	
Nuts	ETS(…)	MAPE	Par. estim.	P-values	Par. estim.	P-value	Par. Estim.	P-value	$\hat{y}$ (000)	APE	$\hat{y}$ (000)	APE
IT	A,A,A	0.37%	0.314	0.001	0.001	0.977	0.001	0.991	22,875	0.1%	23,193	0.8%

Source: ESeC estimates on Eurostat data (Labour Force Survey) – forecast data at 04/08, ex-post at 04/07
Notes: 1. Data for quaterly model assessment from 1999 to 2006 (2007-2008 forecasts). 2. $\hat{y}$: forecast. 3. APE (ex post): $|100\,(y_t - \hat{y}_t)/y_t|$. 4. ETS(A,A,A): Additive Seasonal Smoothing (Winters).

Figure 3: Employment - persons aged 15-64 (thousands), regional (data 1999-06) and national (data 98Q1-07Q3) forecasts, aggregation and conciliation ratio, Italy, 2007-08


	Aggregation (000)		Conciliation ratio
	by Regions	by Quarters	
1999	20,357	…	1.000
2000	20,752	…	1.000
2001	21,169	…	1.000
2002	21,478	…	1.000
2003	21,710	...	1.000
2004	22,060	…	1.000
2005	22,214	…	1.000
2006	22,618	…	1.000
2007	22,941	22,875	0.997
2008	23,275	23,193	0.996

Source: ESeC estimates on Eurostat data (Labour Force Survey).
Notes: Data for model assessment: (i) annual data from 1999 to 2006 (2007-2008 forecasts); (ii) quarterly data from 1998Q1 to 2007Q3 (2007Q4-2008Q4 forecasts). Model: ETS(A,A,A) - Additive Seasonal Smoothing (Winters). Ex-post data represent the observed data not used in the model.

At a European region level the forecast with external conciliation is proportionally adjusted according to the national coefficients (as for the case of Italian regions - see Table 6). In the 2007 APEs are less than 4.5% (8% in 2008). However also the national automatically selected model results in a non-significant set of parameters.

Table 6: Employment - persons aged 15-64 (thousands), regional models, external conciliation forecasts, ex-post APEs by Nuts2, 2007-08

Nuts	$\hat{y}^c$ (000) 2007	APE 2007	$\hat{y}^c$ (000) 2008	APE 2008
ITC1	1,829	0.1%	1,842	0.5%
ITC2	55	0.8%	56	0.3%
ITC3	620	2.3%	624	1.6%
ITC4	4,250	0.4%	4,308	0.8%
ITD1	223	0.4%	225	1.5%
ITD2	215	2.2%	215	3.9%
ITD3	2,084	0.1%	2,110	0.7%
ITD4	510	0.6%	515	0.4%
ITD5	1,885	1.4%	1,907	1.4%
ITE1	1,519	0.3%	1,537	0.4%
ITE2	348	3.4%	353	4.5%
ITE3	640	0.1%	648	0.4%
ITE4	2,120	2.8%	2,152	2.7%
ITF1	499	0.9%	507	0.7%
ITF2	108	2.8%	108	4.2%
ITF3	1,766	3.8%	1,796	8.2%
ITF4	1,276	0.5%	1,311	3.0%
ITF5	195	1.4%	197	2.4%
ITF6	622	4.2%	634	7.5%
ITG1	1,496	1.6%	1,519	3.9%
ITG2	613	1.3%	627	4.2%
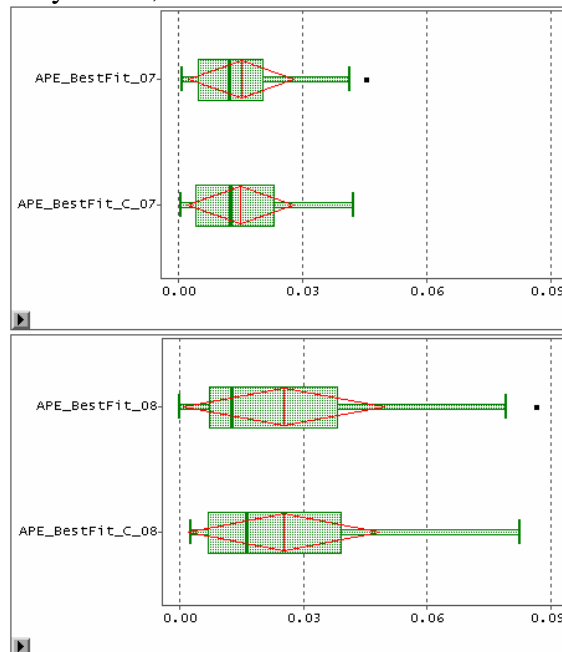
(Box plots: APE_BestFit_07, APE_BestFit_C_07 (axis 0.00–0.09); APE_BestFit_08, APE_BestFit_C_08 (axis 0.00–0.09))

Source: ESeC estimates on Eurostat data (Labour Force Survey).
Notes: 1. Data for annual model assessment from 1999 to 2006 (2007-2008 forecasts). 2. $\hat{y}^c$: conciliated forecasts. 3. APE (ex post): $|100\ (y_t\text{-}^c)/y_t|$; 4. APE_BestFit: APE of estimates (table 4); 5. APE_BestFit_C: APE of conciliated estimates (Table 6).

Model with auxiliary information

In a forecasting framework it is true that the use of exponential weights is a form of prudential behaviour, especially when the time series breakpoint can be detected. In our case the time series can be truncated at this detected point (i.e. 2004Q1) or checked after that breakpoint (i.e. using a dummy as regressor).

The use of the regressor variable or truncated data (with model specified from the remaining data only) can be beneficial to obtain models with a low MAPE and significant parameters (Table 7). It can be noted that if the regional estimates are conciliated with the new national estimates the regional APEs are lower (less than 4.2% in 2007 and less than 7.4% in 2008) (Table 8).

Table 7: Employment - persons aged 15-64 (thousands), national model with truncated data (data 04Q2-07Q3), forecasts, ex-post APE by Nuts0, 2007-08

Nuts	Models ETS(…)	MAPE	Level Par. estim.	Level P-values	Trend Par. estim.	Trend P-value	Seasonal Par. Estim.	Seasonal P-value	$\hat{y}$ (000) 2007	APE 2007	$\hat{y}$ (000) 2008	APE 2008
IT	A,A,A	0.38%	0.078	0.379	0.001	0.995	0.001	0.997	22,863	0.1%	23,144	0.6%

Source: ESeC estimates on Eurostat data (Labour Force Survey) – forecast data at 04/08, ex-post at 04/07
Notes: 1. Data for the assessment of quarterly models from 2004Q2 to 2007Q3 (forecasts from 2007Q4). 2. $\hat{y}$: forecast. 3. APE (ex post): $|100\ (y_t\text{-}\hat{y}_t)/y_t|$. 4. ETS(A,A,A): Additive Seasonal Smoothing (Winters). 5. The break in series of 2004Q1 has been considered.

Table 8: Employment - persons aged 15-64 (thousands), regional models, external conciliation forecasts, ex-post APEs by Nuts2, 2007-08

Nuts	Estimates			
	2007		**2008**	
	$\hat{y}^c$ **(000)**	**APE**	$\hat{y}^c$ **(000)**	**APE**
ITC1	1828	0.1%	1839	0.7%
ITC2	55	1.9%	55	1.7%
ITC3	619	2.4%	623	1.8%
ITC4	4253	0.5%	4310	0.9%
ITD1	223	0.4%	224	1.7%
ITD2	217	1.5%	220	2.0%
ITD3	2083	0.2%	2105	0.9%
ITD4	510	0.7%	514	0.2%
ITD5	1884	1.5%	1903	1.6%
ITE1	1518	0.2%	1533	0.6%
ITE2	347	3.7%	352	4.9%
ITE3	640	0.0%	647	0.2%
ITE4	2119	2.8%	2148	2.9%
ITF1	499	0.8%	506	0.9%
ITF2	108	2.8%	108	4.3%
ITF3	1766	3.8%	1794	8.1%
ITF4	1266	0.3%	1283	0.8%
ITF5	194	1.0%	197	1.9%
ITF6	622	4.2%	633	7.4%
ITG1	1498	1.7%	1521	4.0%
ITG2	615	1.6%	630	4.6%
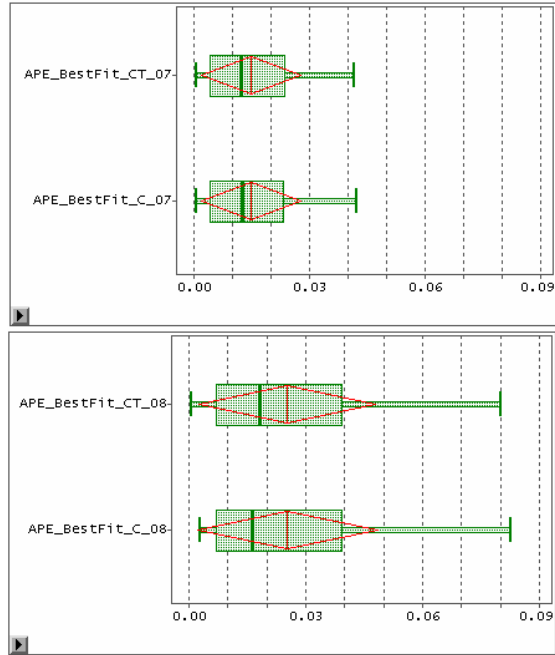
Source: ESeC estimates on Eurostat data (Labour Force Survey).

Notes: 1. Data for annual model assessment from 1999 to 2006 (2007-2008 forecasts). 2. $\hat{y}^c$: conciliated forecasts. 3. APE (ex post): $|100\,(y_t-\hat{y}_t^c)/y_t|$; 4. APE_BestFit_CT: APE Estimates Conciliated to national estimates on truncated data (table 8); 5. APE_BestFit_C: APE of Conciliated Estimates (table 6).

A class of models for regional macroeconomic forecasts

From the classical statistics point of view, while at a national level the auxiliary information can improve the significance of model parameters, at a regional level this is indeed a hard job. However, at a regional level the conciliation process is useful to improve the estimates and to bound the maximum error in the strata. On the other hand, from the economic statistics point of view, the observed macroeconomic data are ill-measured and ill-defined in terms of homogeneity when contextualised in a "historical perspective".

Firstly, these aspects lead to major problems at a regional/sectorial level in terms of:

1. less precision of measurements (i.e. more variability);
2. presence of an exogenous component (i.e. heterogeneity of methods and definitions);
3. less frequent survey periodicity (i.e. absence of a seasonal component);
4. limited history (i.e. difficulties in specifying models).

Secondly, four classes of models can be detected for not seasonal time series:

1. Linear models with homoscedastic errors: ETS(A,N,N), ETS(A,A,N), ETS(A,A$_d$,N);
2. Linear models with heteroscedastic errors: ETS(M,N,N), ETS(M,A,N), ETS(M,A$_d$,N);
3. Multiplicative trend models with homoscedastic errors: ETS(A,M,N), ETS(A,M$_d$,N);
4. Multiplicative trend models with heteroscedastic errors: ETS(M,M,N), ETS(M,M$_d$,N).

Here, we do not consider class 2 and 4 models (i.e. linear models with heteroscedastic errors and multiplicative trend models with heteroscedastic errors) because they give the same points forecast of class 1 and 3 respectively (even if their prediction intervals differ). Thirdly, we do not consider class 3 models (i.e. multiplicative trend models with

homoscedastic errors) because multiplicative trend models are generally not suitable for this kind of macroeconomic annual data. Then, all class 1 models can be written using the following state space equations:

$$y_t = w'x_{t-1} + \varepsilon_t,$$
$$x_t = Fx_{t-1} + g\varepsilon_t,$$

where x_t is the state vector at time t, w and g are column vector, F is a matrix and $\{\varepsilon_t\} \sim$ NID $(0,\sigma^2)$ (having homoscedastic errors, $rx_{t-1} = 1$). Then:

- the ETS(A,N,N) model has $x_t = l_t$, $w = F = 1$ and $g = \alpha$;

- the ETS(A,A$_d$,N) model has $x_t = (l_t, b_t)'$, $w = [1,\phi]$, $F = \begin{bmatrix} 1 & \phi \\ 0 & \phi \end{bmatrix}$ and $g = [\alpha,\beta]'$;

- the ETS(A,A,N) model has the same matrices of ETS(A,A$_d$,N), but with $\phi=1$.

The additive trend method - ETS(A,A,N) - is a special case of damped method obtained letting $\phi=1$ and, if $\beta = 0$, the growth rate is constant over time, and if, in addition, $\alpha = 0$, the level changes at a constant rate over time (the so-called global trend). The growth rate b_t can be positive, negative or zero.

Using the lag operator L and considering stationarity and invertibility conditions, these three models may be also represented as Box-Jenkins models (Box and Jenkins, 1970):

- the ETS(A,N,N) model may be represented as an ARIMA(0,1,1) model;
- the ETS(A,A$_d$,N) model may be represented as an ARIMA(1,1,2) model;
- the ETS(A,A,N) model may be represented as an ARIMA(0,2,2) model.

Taking into proper account the nature of the data and all the above considerations, the linear damped model with homoscedastic errors may be used as the prudential reference model for forecasting. From a practical point of view, regional "Damped" models present conciliated APEs (Figure 3) which are of the same magnitude of the "Best-Fit" models APEs (i.e. slightly more than 4% in 2007 and slightly more than 8% in 2008). These models, further to provide estimates without trend components and since they are less parsimonious in terms of parameters with respect to ETS(A,N,N) (APEs less than 2% are observed for ITC2 and ITD2), tend to downsize the growth (or fall) rates effect when these rates are thought not to prevail over time (e.g. cyclic factors, market saturation, etc.). For example, in 2008 the median of the "Damped" model APEs is less than the "Best-Fit" APEs (1.74% vs. 1.83%), whereas the maximum APE is higher, due to the conciliation effect (i.e. ITF3 already was a damping model).

Figure 3: Employment - persons aged 15-64 – Damped ES Model, regional ex-post APEs by Nuts2, 2007 and 2008
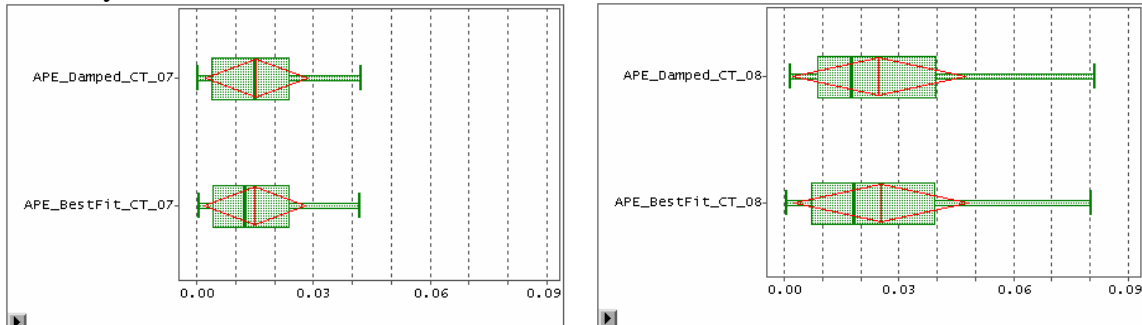
Source: ESeC estimates on Eurostat data (Labour Force Survey). **Notes**: 1. APE_BestFit_CT: APE Estimates Conciliated to national estimates on truncated data (table 8); 2. APE_Damped_CT: APE Estimates of ETS(A, A$_d$, N) conciliated to national estimates on truncated data.

Conclusions

When a few information is available, in order to get reliable estimates it can be necessary to use databases other than those for the specific analysis at hand. The management of multiple sources of data can improve forecasts, as the herein presented results have shown (e.g. the 2007 national APE from regional aggregation is 0.4%, whereas the APE from regional aggregation after reconciliation is 0.1%). National data are often provided before the regional data, so that, as explained in the case of the employment level, the proposed procedure allows for a reliable estimate of regional data in advance of its publication by offices of national statistics.

References

Box G.E.P., Jenkins G.M. (1970) *Time series analysis: forecasting and control*, Holden-Day, San Francisco.

Brown R.G. (1959) *Statistical forecasting for inventory control*, McGraw-Hill, New York.

Chatfield C., Koehler A.B., Ord J.K., Snyder R.D. (2001) A new look at models for exponential smoothing, *The Statistician*, **50**, Part 2, 147-159.

Granger C.W.J., Newbold P. (1986) *Forecasting economic time series*, Academic Press, USA.

Holt C.C. (1957) Forecasting trends and seasonals by exponentially weighted averages, O.N.R. Memorandum 52/1957, Carniage Institute of Technology.

Hyndman R.J., Koehler A.B., Ord J.K, Snyder R.D. (2008) *Forecasting with exponential smoothing*, Springer-Verlag, Berlin.

Marcellino M. (2004) Forecasting EMU macroeconomic variables, International Journal of Forecasting, 20, 359-372.

Santamaria L. (2000) *Analisi delle serie storiche economiche*, Vita e Pensiero, Milano, Italy.

Taylor J.W. (2003) Exponential smoothing with a damped multiplicative trend, *International Journal of Forecasting*, 19, 715-725.

Verrecchia F. (2008) Previsione e selezione automatica dei modelli per serie storiche regionali: metodo bi-fase a conciliazione esterna, in: SAS Business Analytics Gallery 2008, Rome. [Poster]. Online: Available [http://www.economicstatistics.eu/poster].