A monthly indicator of employment in the euro area: real time analysis of indirect estimates

Moauro, Filippo

30 December 2010

A monthly indicator of employment in the euro area: real time analysis of indirect estimates*

Filippo Moauro†

December 30, 2010

Abstract

The paper presents the results of an extensive real time analysis of alternative model-based approaches to derive a monthly indicator of employment for the euro area. In the experiment the Eurostat quarterly national accounts series of employment is temporally disaggregated using the information coming from the monthly series of unemployment. The strategy benefits of the contribution of the information set of the euro area and its 6 larger member states, as well as the split into the 6 sections of economic activity. The models under comparison include univariate regressions of the Chow and Lin' type where the euro area aggregate is directly and indirectly derived, as well as multivariate structural time series models of small and medium size. The specification in logarithms is also systematically assessed. The largest multivariate setups, up to 49 series, are estimated through the EM algorithm. Main conclusions are the following: mean revision errors of disaggregated estimates of employment are overall small; a gain is obtained when the model strategy takes into account the information by both sector and member state; the largest multivariate setups outperforms those of small size and the strategies based on classical disaggregation methods.

Keywords: temporal disaggregation methods, multivariate structural time series models, mixed-frequency models, EM algorithm, Kalman filter and smoother

JEL classification: C22, C32, C51, C52

*The views expressed in this paper are those of the author and do not necessarely reflect the positions of policies of Eurostat

†Eurostat, 5 Rue A. Wecker, L-2721 Luxembourg, E-Mail: Filippo.Moauro@ec.europa.eu

1 Introduction

The need to provide official short term statistics has increased recently for requests coming from several directions at national or international level. With this purpose Eurostat disseminates a wide spectrum of monthly and quarterly key set of indicators which provide a comprehensive picture of the macroeconomic development in the euro area and the European Union. We refer to the Euro-IND database, which covers all the most important domains of macroeconomic measures, among which labour market statistics of high professional standards (`http://epp.eurostat.ec.europa.eu/portal/page/portal/euroindicators/labour_market/database`). The latter contains, amongst others, time series of employment, unemployment, labour inputs, hours worked and labour cost indexes.

Despite the substantial improvements in length, coverage and timeliness of labour market indicators for the euro area, national account measures of employment are available only quarterly, whereas it would be of sure interest to have corresponding estimates of employment at monthly frequency.

In recent periods, characterized by higher uncertainty on the effective exit from the recent recession, employment represents a leading macroeconomic variable. The analysis of labour markets represents a crucial element for policy makers, economic analysts, researchers and business cycle experts. The relevance of this variable and the possibility to get a monthly estimate of employment motivate this work. In particular we advocate the use of an indirect approach to produce monthly estimates of euro area employment through advanced statistical methodology and the availability of reliable and informative related indicators. The alternative of new statistical surveys for monthly measures of employment is rarely sustainable from most statistical institutions because of the high costs for their implementation.

Temporal disaggregation methods play an important role to derive relevant short term statistics like quarterly accounts or unemployment data. The most popular techniques is based on the work by Chow & Lin (1971) and its extensions by Fernández (1981) and Litterman (1983). All these methods consist in regressions with autocorrelated residuals, where the information coming from related quarterly indicators is used for distributing the available series of annual or quarterly totals. The method is easy to implement and results of disaggregation are characterized by an high stability and robustness. More recently a unified state space approach has been proposed by Proietti (2006*b*) for dynamic regression methods that depend on a single autoregressive parameter. Further extensions move towards Seemingly Unrelated Time Series Equations (SUTSE) models for temporal disaggregation, like in Harvey & Chung (2000) and Moauro & Savio (2005). In these works the target series to disaggregate and related indicators are all treated as dependent variables in a Mixed-Frequency (MF) SUTSE model. Further, Proietti (2006*a*) extends temporal disaggregation to data subject to a Box-Cox transformation.

On the empirical side Proietti & Moauro (2006) has recently proposed a solution to the problem of estimating a monthly GDP series for the US and the euro area economy using data with different frequency of observation within the Stock & Watson (1991) dynamic factor model. MF-dynamic factor models have been also developed by Frale et al. (2010*a*) and Frale et al. (2010*b*) for the construction of a monthly indicator of GDP for the euro area by components of the supply and the demand side. Finally in Proietti (2008) the euro area monthly aggregates are estimated along with the corresponding components of the largest member states, i.e. Germany, France, Italy, Spain, Belgium and the Netherlands within a model of large size.

This paper aims at assessing a set of alternative model-based approaches to derive a monthly indicator of employment in the euro area through an extensive real time analysis. In the experiment the Eurostat quarterly national accounts series of employment is temporally disaggregated using monthly unemployment. The strategy benefits of the contribution of the information set of the euro area and its 6 larger member states (MS), as well as the split of the economy into the 6 NACE sections of economic activity. The models under comparison include univariate regressions of the Chow and Lin' type where the euro area aggregate is directly and indirectly derived, as well as SUTSE models (see, e.g., Harvey (1989), pp. 429-435; Fernández & Harvey (1990); Harvey & Koopman (1997)) of small and medium size. The specification in logarithms is also systematically assessed. Estimation of SUTSE models is carried out through the EM algorithm.

Main original features of the paper are the following: first, it represents, at our knowledge, the first systematic work on monthly measures of employment in the euro area; second, it offers a wide spectrum of model strategies, from those where the entire dataset is splitted into smaller sub-datasets according to economic considerations, to a single SUTSE model for the entire dataset; third, it implements an extensive real time experiment using the Eurostat vintage database to access the reliability of monthly estimates and to compare the results of alternative model strategies; fourth, all the strategies employed in the experiment provide explicit measures of uncertainty around final monthly estimates; finally, it is proposed an iterative scheme based on the EM algorithm provided by Koopman (1993) for the statistical treatment of MF-SUTSE models,

The paper is organized as follows: Section 2 introduces temporal disaggregation methods used in the experiment; Section 3 discusses strategies and results of model estimation; Section 4 the comparative real time analysis, whereas Section 5 shortly concludes. Two appendixes are provided for state space forms of linear and non-linear SUTSE temporal disaggregation and the EM algorithm.

2 Methods used in the experiment

Several are the model-based strategies of temporal disaggregation. In the experiment here developed for the euro area employment, no a-priori choice has been adopted. The aim is a fair comparison between simple regression methods of the Chow-Lin type commonly used among statistical agencies and the more complex formulations of the SUTSE modelling class. In the former case the advantage is the user-friendliness of the application, whereas in this latter the possibility of setting up models of small and medium size, splitting the economy in sections of economic activities and/or including MS data. The comparison concerns also the model formulation in logarithms.

2.1 Regression methods

Regression methods of temporal disaggregation hypothesize a simple linear univariate relationship between the dependent variable y_t and a set of related indicators x_t. y_t is available in form of sum or average over a given interval of observation (e.g. every quarter), whereas the covariates x_t are available at a higher frequency (e.g. every month). A general setup is the Autoregressive Distributed Lag model ADL(1,1) formulated at the higher frequency:

$$\Delta^l y_t = \rho\Delta^l y_{t-1} + m + gt + \beta_0'\Delta^l x_t + \beta_1'\Delta^l x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim NID\left(0, \sigma^2\right), \quad (1)$$

where Δ is the difference operator such that $\Delta y_t = y_t$ - y_{t-1}, y_t is a scalar, l is for the order of differentiation which in our applications takes the values 0 or 1 only, ρ is the autoregressive term such that $-1 < \rho < 1$, m is a constant term, g is the coefficient associated to the trend; the vectors β_0 and β_1 are for the coefficients of related indicators x_t modelled respectively at lag 0 and 1; ε_t is the residual white noise with variance σ^2.

The ADL(1,1) model (1) nests, under suitable assumptions about initial conditions, the forms developed by Chow and Lin, Fernandez and Litterman (denoted respectively CL, FE and LT henceforth). In particular the CL model occurs when $l = 0$, i.e. the series is modeled in levels and $\beta_1 = -\rho\beta_0$. In this case the model (1) becomes a stationary regression model with AR(1) residuals. The FE model when $l = 1$, $\rho = 0$ and $\beta_1 = 0$, or l

= 0, ρ = 1 and $\beta_1 = -\beta_0$, implying non-stationary I(1) residuals; finally, the LT model occurs when $l = 1$ and $\beta_1 = -\rho\beta_0$, i.e. a regression with non-stationary ARI(1,1) residuals. Further forms employed in the comparison are two ADL(1,0) models, which imply $\beta_1 = 0$ and $l = 0$ or $l = 1$, respectively if the series are in levels or in first differences.

Statistical treatment of model (1) is quite convenient since all the regression coefficients m, g, β_0, β_1 and the variance term σ^2 could be concentrated out from the log-likelihood function, which becomes a function of the parameter ρ alone. Then, maximum likelihood estimation is conveniently carried out via grid search over the interval (-1, 1). When y_t is subject to logarithmic transformation, estimation require a nonlinear statistical treatment, since temporal disaggregation becomes a nonlinear problem. Here we adopt the unified state space approach provided by Proietti (2006*a*).

2.2 SUTSE models

SUTSE models are widely treated in the literature and represent a multivariate generalization of structural time series models (see, e.g., Harvey (1989), pp. 429-435; Fernández & Harvey (1990); Harvey & Koopman (1997)). Given a cross-section of time series $y_t = (y_{1t}, \ldots, y_{Nt})'$, it is assumed that each y_{it}, $i = 1, 2, \ldots, N$ and $t = 1, 2, ..., n$, is not directly related with the others, although the series are subject to similar influences. y_t is expressed in terms of additive N-dimensional unobserved components, e.g. level μ_t, slope β_t, and irregular ξ_t, which can be contemporaneously correlated. Here a multivariate local linear trend (LLT hereafter) model is considered, where $\mathbf{y}_t$ consists of a stochastic trend plus a white noise:

$$\begin{aligned} y_t &= \mu_t + \xi_t, & \xi_t &\sim NID(0, \Sigma_\xi), & (2)\\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, & \eta_t &\sim NID(0, \Sigma_\eta), & (3)\\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim NID(0, \Sigma_\zeta). & (4)\end{aligned}$$

The LLT model (2)-(4) have a standard statistical treatment within the Kalman filter (KF) when y_t is fully observed (for details look at the STAMP manual by Koopman et al. (2009)). When the components $y_{1t}, \ldots, y_{Nt}$ are available at different frequencies and there is the need of temporal disaggregation, the KF is also a viable tool adopting the approach by Harvey & Chung (2000) when data are in levels and by Proietti & Moauro (2006) when data are in logs. Complexity in estimation concerns only the number of hiperparameters pertaining the 3 disturbance covariance matrices Σ_ξ, Σ_η, and Σ_ζ which increases with the dimension of the system. When N is large the KF is still an efficient instrument for log-likelihood evaluation even if computationally demanding, but maximization is no more feasible through gradient based methods. An alternative solution is provided by the EM algorithm of Koopman (1993). Here this method is applied, developing an extension to handle linear and non-linear temporal disaggregation under mixed frequency (MF) SUTSE models. The SSF of the LLT model (2)-(4) is presented in the appendix together with details on the statistical treatment of temporal disaggregation and the EM algorithm implemented in the application. Note that in this paper the diffuse KFS due to de Jong (1991) is adopted for the treatment of initial conditions.

3 Strategy and results of model estimation

This section presents the strategy and results of model estimation to derive a monthly indicator of the euro area employment for the total economy, as well as the 6 sections of the NACE Rev.1.1 classification of economic activities. Both direct and indirect strategies are followed: in the former total quarterly employment is modelled directly with the related indicator of unemployment; in the latter a model is fitted to sectoral data and the total is derived by summing up the monthly estimated subtotals. Under the SUTSE modelling approach it is also assessed the contribution of larger MS data. Finally, a mixed-frequency SUTSE model of medium size including 49 series is fitted simultaneously to both the euro area and MS data of unemployment and employment sectoral components.

The underlying idea in the use of sectoral data is to follow a strategy which embodies and translates the standard practice of NSIs of compiling official statistics from a detailed set of coherent and integrated measures of economic activities. The application is novel to my knowledge and challenging in the sense that data are available mainly quarterly at both the euro area level and by MS, the time series cover a short sample period and related monthly indicators are quite a few.

3.1 The dataset

The series used in the experiment are in total 56, all are seasonally adjusted and represented in Figures 1-7 respectively for the euro area of 12 MS (Germany, France, Italy, Spain, The Netherlands, Belgium, Luxembourg, Portugal, Austria, Greece, Finland and Ireland), plus Belgium, Germany, Spain, France, Italy and the Netherlands. The plots represent the levels of the series at the monthly frequency. Below is given a short presentation of the series by variable:

1. the quarterly national accounts (QNA) series of employment available from the first quarter of 1995 to the third quarter of 2009 released by Eurostat in January 2010. In total 49 series including 7 series of employment for the total economy and the 6 components of the A6 breakdown: 1.Agriculture, hunting, forestry and fishing (AGR), 2.Industry (IND); 3.Construction (COS); 4.Trade, hotels and restaurants, transport and communication (TTC); 5.Financial services and business activities (FBS); 6)Other services (OTS);

2. the monthly series of total unemployment in thousands of persons for all the same geographical entities in the sample January 1995-December 2010 released by Eurostat in February 2010;

Note that we do not use in the experiment monthly Labour input indexes (source Eurostat), albeit these data are also related to employment. Two are the reasons: first these data give a partial representation of total economy since only 8 indexes are available for Industry (euro area, Belgium and Germany), Construction (euro area and Germany) and Trade (euro area, Germany and Spain); second, there is a need to conform the setup to the real time experiment of next section, where we employ 73 monthly releases for unemployment and 45 for employment starting from the release of February 2004; by contrast, the Eurostat vintage database in use here includes only 11 releases for labor inputs staring from that of July 2009. Furthermore, the hours worked series released by Eurostat do not appear of utility in this context since extremely erratic.

3.2 Estimation results of regression methods

Table 1 and 2 present the results of model estimation for temporal disaggregation methods based on regressions of model (1). The results refer to direct estimation of euro area employment for the total economy (y_t), using monthly unemployment as a unique related indicator (x_t). In total 7 specifications have been fitted to data in the sample period from January 1995 to December 2009: from the CL, FE and LT models, to the ADL(1,0) and ADL(1,1) setups both in levels and first differences; the latter are denoted respectively as ADL(1,0)D and ADL(1,1)D. Two parallel setups are implemented: in the former, results in table 1, the formulations in levels and in the latter those in logarithms ($\times 100$), whose results are presented in table 2. Note that, model estimates of the indirect experiments, not presented here for reason of space, are available upon request.

From left to right both the tables present the maximised profile likelihood $\widehat{\mathcal{L}}$, the autoregressive coefficient $\widehat{\rho}$, the constant term $\widehat{c}$, the trend coefficient

$\widehat{g}$ and the coefficients $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for unemployment, modelled respectively at lag 0 and 1; t-student statistics, in parenthesis, are presented below each regression coefficients.

From table 1 it emerges the presence of constant and trend in all the model specifications apart ADL(1,0)D. The estimated autoregressive term $\widehat{\rho}$ is always positive and in most cases close to 1, which would imply non-stationarity in model residuals. The estimates of other regression coefficients are all significant with the exception of $\widehat{g}$ for ADL(1,1)D, whose t-student statistic is however -1.75. As expected the estimated coefficient $\widehat{\beta}_0$ is negative for all specifications, since employment is inversely related to unemployment. The only exception is ADL(1,0)D where, however, unemployment doesn't seem to fit appropriately to employment given the estimate of $\widehat{\beta}_0$ close to nil.

Further, note in table 1 the value $\widehat{\rho} = .05$ for LT: it implies that the fit is similar to that of the FE model, as it is clear from the identical value of $\widehat{\mathcal{L}}$ $= -450.42$ and from estimated regression coefficients $\widehat{c}$, $\widehat{g}$ and $\widehat{\beta}_0$ very close to each others. Finally, the estimate of the more complex model ADL(1,1)D produces an estimate $\widehat{\rho}$ almost nil, which would imply over-parametrization and non-stationarity for this particular setup.

Similar considerations come from specifications in logarithms: from table 2 it appears once again confirmed the negative relation employment-unemployment, since estimated coefficients $\widehat{\beta}_0$ are always negative for all model specifications, apart the case of ADL(1,0)D, as discussed for models in levels; the estimates $\widehat{\rho}$ are similar to the corresponding models in the levels. By contrast the trend coefficient $\widehat{g}$ is either close to nil like in the CL and ADL(1,1) specifications, or completely insignificant like in the other 4 cases. In other words, when data are in logs the model (1) fitted to employment does not require the addition of a deterministic trend to unemployment and, in this respect models in logs should be preferred to those in levels.

3.3 Estimation results of SUTSE models

The implementation of the SUTSE approach here discussed refers to the simultaneous model of euro area and MS data with 49 time series: notably 42 quarterly series of employment by MS and by sector of economic activity, plus the 7 related monthly indicators of total unemployment. The monthly euro area and MS totals are obtained by summing up the 6 corresponding estimates of sectoral components. The sample is January 1995-December 2009 and the model is specified in logarithms.

Results of the simultaneous approach in levels and of simpler direct and indirect procedures are not presented here for reason of space, but are available upon request.

In model estimation two restrictions have been considered: the former consists in assuming a constant slope in the LLT model, which means in (4) to restrict to nil the covariance matrix Σ_ζ associated to the slope component β. As a consequence the series are imposed to be $I(1)$ with trend given by a random walk with drift; the latter that the irregular component is taken to be idiosyncratic, corresponding to a diagonal covariance matrix Σ_ξ in (2).

The estimation of the LLT model and temporal disaggregation is carried out using the EM algorithm presented in appendix 2. As starting values for Σ_η and Σ_ξ it have been considered, respectively, two diagonal matrices, with elements provided by the corresponding disturbance variances estimated from a univariate setup applied to all the 49 single series. Convergence of EM iterations occurs when the system matrices Σ_η and Σ_ξ and disaggregated estimates do not differ from one iteration to another by more then a specified tolerance, in our case fixed to 10^{-5}.

The EM algorithm required 88 iterations for convergence. The profile log-likelihood has been equal to -14269.3. Tables 3 presents parameter estimates of standard deviations of disturbances for the level and the irregular components obtained computing the root-square of diagonal elements of $\widehat{\Sigma}_\eta$ and $\widehat{\Sigma}_\xi$ respectively. For the 42 series of employment it emerges a general equilibrium between level and irregular components since estimated standard deviations of disturbances show similar size. By contrast, variation of trend dominates that of irregular component for unemployment.

From the estimated matrices Σ_η, a set of 49 correlations are computed between the level component disturbances of each series and that of the euro area total employment. These values are presented in table 4, from which it emerges that both size and sign of correlations are generally in line with expected results: in absolute terms higher correlations of total employment emerge with the corresponding euro area sectoral measures and with euro area unemployment but with a negative sign. Figure 8 displays temporal disaggregated estimates of employment by sector of economic activity in the euro area and figure 9 of total employment by MS.

4 Comparative real time analysis

The quality of disaggregated estimates relative to the strategies described in previous section have been assessed by looking real time at their revision histories. The exercise is based on the vintages of employment and unemployment data available at Eurostat from February 2004 to January 2010.

The revision histories has been generated as follows: the start is February 2004 the 3-rd, when the monthly series of unemployment is released from January 1995 to December 2003 and employment is available from 1995 1-st quarter to 2003 3-rd quarter; based on this information set a first monthly estimate is performed for the sample January 1995-December 2003. At the beginning of March 2004 unemployment data are released for January 2004 and the model is re-estimated over a sample period augmented by 1 observation; accordingly, the employment release used at this stage is that available at this date. Iterating on the same experiment every month until February 2010 the 1-st, it produces a triangle of 73 monthly estimates. From these data quarterly totals and revision errors are computed.

Figure 10 reports monthly estimates of employment in the period January 1995 - December 2009 obtained running the contemporaneous SUTSTE model for the 73 vintages considered in the experiment. More relevant changes in the estimates occur when new quarterly employment observations are released. Overall, real time estimates show a substantial degree of reliability. Focusing in the months of 2008-2009 it is visible the effect of the recent recession on the monthly estimates, the decreasing pattern and the uncertainty on the end of the recession.

Tables 5 and 6 resume the results of the real time experiment for model specifications in logarithms, comparing root mean squared revision errors (RMSRE) of quarterly growth rates of sectoral euro area employment. Tables 5 is for regression methods of section 2.1 and table 6 for SUTSE models of section 2.2. The estimates for total economy refer to both the direct and indirect approaches. In bold are the lowest sectoral RMSREs within the same table and with an asterisk is denoted the lowest among the two classes of models.

Table 6 presents RMSRE statistics related to four SUTSE modelling approaches: a) indirect approach implying 6 separate model specifications by sector with the contribution of MS data; b) direct approach, i.e. same strategy as a) but on euro area and MS data for the total economy; c) contemporaneous specification of the 6 sectors without MS data; d) contemporaneous specification with the full split of data by sector and MS (that of model estimates presented in previous section).

Specifications in logarithms of tables 5 and 6 are exhaustive of the entire real time experiment, since in all cases they outperform correspondent models in the levels. Looking at the last column of both tables 5 and 6 it emerges that RMSREs of quarterly growth rates are in the range between 0.25-0.29. Higher values result when the comparison concerns sectoral estimates: in agriculture (AGR) the range of RMSREs is 0.74-0.90, in industry (IND) 0.41-0.54, in construction (COS) 1.05-1.16, in trade, transport and communication

(TTC) 0.47-0.51, in financial and business services (FBS) 0.47-0.51 and in other services (OTS) 0.39-0.42.

From a general comparison between the two tables it emerges that SUTSE models outperform regression methods in 4 out of the 7 groupings of economic activity considered in the example. Moreover the contemporaneous SUTSE modelling approach with full split of data outperforms all the other cases for the total economy, IND and FBS, with consistent gains in terms of RMSREs. By contrast, regression methods outperforms SUTSE models in the two sectors characterized by higher revision errors like Agriculture and Construction.

Overall, what the real time analysis here implemented seems to suggest is that the use of larger modelling setups like those including MS data and the split into sectors of economic activity could increase accuracy in constructing euro area monthly indicator of employment. The exercise also indicates the effectiveness of nonlinear disaggregation methods for the treatment of data subject to logarithmic transformation and the use of the EM algorithm for model estimation.

5 Conclusions

The paper discusses temporal disaggregation methods and model strategies to derive a euro area monthly indicator of employment. The main contribution is to provide and compare a wide spectrum of monthly disaggregated estimates that are consistent with available quarterly totals. Most interesting alternatives are modelled in logarithms and benefit from one side of the contribution of the information set of 6 largest member states data and from the other of the split among sections of economic activity.

The study is complemented by an extensive real time analysis using the Eurostat vintage database of employment and unemployment, which provided an average measure of revisions of monthly disaggregated estimates.

From a computation standpoint the application indicates how the EM algorithm could open the accessibility to models of medium scale even for problems of estimation which require high stability and robustness, like those employed for the intensive production of official statistics. This can be particularly effective in increasing the efficiency of estimation and in reducing revisions of released official statistics.

Appendix 1: State space form for SUTSE temporal disaggregation

Let's denote the 'model frequency' of the system (2)-(4) with δ and the frequencies at which the unobserved disaggregated flows are observed with $\delta_1^\dagger, \ldots, \delta_N^\dagger$. Model and observed frequencies are such that their ratios, denoted $\delta_i = \delta/\delta_i^\dagger$, are integers for each i. Then a cumulator variable y_t^c is introduced, such that

$$\begin{aligned} y_t^c &= C_t y_{t-1}^c + y_t = C_t y_{t-1}^c + \mu_t + \xi_t = \\ &= C_t y_{t-1}^c + \mu_{t-1} + \beta_{t-1} + \eta_t + \xi_t, \end{aligned} \tag{5}$$

where $C_t = diag\,(c_{1t}, c_{2t}, \ldots, c_{Nt})$ and:

$$c_{it} = \begin{cases} 0 & t = 1, \delta_i + 1, 2\delta_i + 1, \ldots \\ 1 & \text{otherwise.} \end{cases} . \tag{6}$$

The need of the cumulator emerges for the component series in y_t subject to temporal aggregation, otherwise $c_{it} = 0$ for $t = 1$, ..., n and $y_{it}^c = y_{it}$. Equations (2)-(4) allow a compact state space form (SSF hereafter). Defining the state vector like $\alpha_t = (\mu_t', \beta_t', y_t^{c\prime})'$ the SSF is:

$$\begin{aligned} y_t &= Z\alpha_t, \ t = 1, ..., n, & (7)\\ \alpha_t &= T_t \alpha_{t-1} + H\varepsilon_t, \ t = 2, ..., n, & (8)\\ \alpha_1 &= W\gamma + H\varepsilon_1, & (9) \end{aligned}$$

$$\varepsilon_t = (\eta_t', \zeta_t', \xi_t')', \quad \varepsilon_t \sim NID(0, \sigma^2 I) \tag{10}$$

$$Z = [0, 0, I_N], \quad \gamma = (\mu_0', \beta_0')', \tag{11}$$

$$T_t = \begin{pmatrix} I_N & I_N & 0 \\ 0 & I_N & 0 \\ I_N & I_N & C_t \end{pmatrix}, \quad H = \begin{pmatrix} \Gamma_\eta & 0 & 0 \\ 0 & \Gamma_\zeta & 0 \\ \Gamma_\eta & 0 & \Gamma_\xi \end{pmatrix}, \quad W = \begin{pmatrix} I_N & I_N \\ 0 & I_N \\ I_N & I_N \end{pmatrix}, \tag{12}$$

where the system matrix T_t is time varying for the presence of the matrix C_t, the matrix H contains the blocks Γ_h's derived through Cholesky decomposition of Σ_h such that $\Sigma_h = \Gamma_h \Gamma_h'$, $h = \eta$, ζ, ξ. In the equation for the initial state (9) the vector of regression coefficients γ simply includes the level and slope components μ_0 and β_0 at time 0.

Consider now the case when the time series into y_t are expressed in logarithm: then it is useful to consider a partition of y_t such that $y_t = (y_{1t}', y_{2t}')'$, where the first block y_{1t} includes the fully observed subset of N_1 data and y_{2t} the block of N_2 elements subject to temporal aggregation, with $N = N_1 + N_2$. In this case the aggregation constrain is linear in $Y_{2t} = \exp(y_{2t})$ since the aggregated series are given by

$$Y_{2\tau}^{*}=\sum_{i=0}^{\delta-1}Y_{2,\tau\delta-i},\quad \tau=1,...,[n/\delta]\,, \tag{13}$$

where $[n/\delta]$ is for the integer part of n/δ. Following the discussion in Proietti & Moauro (2006), a non-linear SSF under the temporal aggregation constraint (13) is required. The new cumulator variable Y_{2t}^{c} to be included in a new observational vector $Y_t^{\dagger}$ is such that

$$\begin{array}{ll} Y_t^{\dagger}=(y_{1t}',Y_{2t}^{c\prime})', & C_t=diag\,(0,C_{2t})\,, \\ Y_{2t}^{c}=C_{2t}Y_{2,t-1}^{c}+\exp(y_{2t})\,, & t=1,...,n, \end{array} \tag{14}$$

where the diagonal matrix C_{2t} is for selecting from y_t only the time series subject to the non-linear temporal aggregation (13). Note that the first block y_{1t} of $Y_t^{\dagger}$ is in logarithm and the second in levels, represented by the cumulator variable Y_{2t}^{c}. This latter variable is observed only at time δ, 2δ, ... The new variable $Y_t^{\dagger}$ is included into a new state vector $\alpha_t^{\dagger}=\left(\mu_t',\beta_t',Y_t^{\dagger\prime}\right)'$ around which a linear and Gaussian approximating (LGA) model could be developed.

The statistical treatment of the LGA model is carried out through approximate conditional mode estimation by extended KFS based on , Durbin & Koopman (2001) and Fahrmeir (1992). Given an arbitrary trial value $\widetilde{y}_{2t}$ of y_{2t} the LGA model is obtained from the first-order Taylor series expansion of the cumulator around this value:

$$\begin{aligned} Y_{2t}^{c} &= C_{2t}Y_{2,t-1}^{c}+\exp(\widetilde{y}_{2t})+\widetilde{D}_{2t}\,(y_{2t}-\widetilde{y}_{2t}) \\ &= C_{2t}Y_{2,t-1}^{c}+\left[\exp(\widetilde{y}_{2t})-\widetilde{D}_{2t}\widetilde{y}_{2t}\right]+\widetilde{D}_{2t}\,(\mu_{2t}+\xi_{2t}) \\ &= C_{2t}Y_{2,t-1}^{c}+\left[\exp(\widetilde{y}_{2t})-\widetilde{D}_{2t}\widetilde{y}_{2t}\right]+\widetilde{D}_{2t}\left(\mu_{2,t-1}+\beta_{2,t-1}\right)+\widetilde{D}_{2t}\,(\eta_{2t}+\xi_{2t})\,, \end{aligned}$$

where $\widetilde{D}_{2t}=diag\,[\exp(\widetilde{y}_{2t})]$, i.e. the matrix of derivatives of each element of the cumulator Y_{2t}^{c} with respect to y_{2t}, evaluated at the trial value $\widetilde{y}_{2t}$ and the vectors μ_{2t}, β_{2t}, η_{2t} and ξ_{2t} result, respectively, by the partition of the corresponding vectors μ_t, β_t, η_{2t} and ξ_t such that

$$\mu_t=(\mu_{1t}',\mu_{2t}')',\quad \beta_t=(\beta_{1t}',\beta_{2t}')',\quad \eta_t=(\eta_{1t}',\eta_{2t}')',\quad \xi_t=(\xi_{1t}',\xi_{2t}')'.$$

The SSF of the LGA model is given by

$$
\begin{aligned}
Y_t^{\dagger} &= Z\alpha_t^{\dagger},\ t=1,...,n, && (15)\\
\alpha_t^{\dagger} &= T_t^{\dagger}\alpha_{t-1}^{\dagger}+c_t+H_t^{\dagger}\varepsilon_t,\ t=2,...,n, && (16)\\
\alpha_1^{\dagger} &= W^{\dagger}\gamma+c_1+H_1^{\dagger}\varepsilon_1, && (17)
\end{aligned}
$$

with the definitions for Z and γ equal to those of equations (11) and (12) and the system matrices $T_t^{\dagger}$, c_t, $W^{\dagger}$, and $H_t^{\dagger}$ like

$$
\begin{array}{ll}
T_t^{\dagger}=\begin{pmatrix} I_N & I_N & 0\\ 0 & I_N & 0\\ \widetilde{D}_t & \widetilde{D}_t & C_t\end{pmatrix}, & c_t=\begin{bmatrix} 0\\ \exp\left(\widetilde{y}_{2t}\right)-\widetilde{D}_{2t}\widetilde{y}_{2t}\end{bmatrix},\\
W^{\dagger}=B_1W, & H_t^{\dagger}=B_tH;
\end{array} \qquad (18)
$$

the matrices W and H are like in equation (12) and the time-varying matrices $\widetilde{D}_t$, and B_t are such that

$$
\widetilde{D}_t=\begin{pmatrix} I_{N_1} & 0\\ 0 & \widetilde{D}_{2t}\end{pmatrix},\quad B_t=diag\left(I_{2N},\widetilde{D}_t\right).
$$

Note that the disturbance vector ε_t is such that $\varepsilon_t=\left(\eta_t',\zeta_t',\xi_t'\right)'$ like in equation (10) albeit, in this case, it is referred to the logarithms of observed data. The covariance matrix of ε_t is here denoted as Ω and it is such that $\Omega=HH'$. The time varying elements into B_t derive from the Taylor expansion of the non-linear cumulator around to the trial value $\widetilde{y}_{2t}$.

Appendix 2: The EM algorithm for linear and non-linear temporal disaggregation

Linear aggregation constraint: The derivation of the EM algorithm for SUTSE models finds a effective solution since unknown parameters are found exclusively in the covariance matrices of the state space form (see Koopman (1993)). Let's denote as $\widetilde{\Omega}=\widetilde{H}\widetilde{H}'$ an initial trial estimate of the covariance matrix for the SSF (7)-(12). Following Koopman (1993), the EM step gives a new covariance matrix $\hat{\Omega}$ which always increases the log-likelihood. The new covariance matrix $\hat{\Omega}$ is such that

$$
\hat{\Omega}=\widetilde{\Omega}+\widetilde{\Omega}\Theta_r\widetilde{\Omega},\quad \Theta_r=1/n\sum_{t=1}^{n}\left(\hat{r}_t\hat{r}_t'-\hat{N}_t\right), \qquad (19)
$$

where $\hat{r}_t$ is the $m-vector$ of disturbance estimates at time t conditional to the information in the whole sample and $\hat{N}_t$ the corresponding $(m \times m)$ covariance matrix. These quantities are obtained through the Koopman's (1993) disturbance smoother.

Non-linear aggregation constraint: Albeit the LGA model (15)-(18) induces a time-varying SSF, unknown parameters of the model remain into the covariance matrix of disturbances. Then, the updating equation (19) is simply modified in order to include the time-varying elements of B_t like in the expressions below

$$\hat{\Omega}=\widetilde{\Omega} + \widetilde{\Omega}\Theta_r^\dagger\widetilde{\Omega}, \quad \Theta_r^\dagger = 1/n \sum_{t=1}^{n} B_t^{-1} \left(\hat{r}_t\hat{r}_t' - \hat{N}_t\right) B_t^{-1}. \tag{20}$$

The proof is straightforward following Koopman (1993) and Durbin & Koopman (2001).

Acknowledgements

A preliminary version of the paper was presented at the 6th Eurostat Colloquium on Modern Tools for Business Cycle Analysis: the lessons from global economic crisis, Luxembourg, 26-29 September, 2010 and at the 4th CSDA International Conference on Computational and Financial Econometrics, London, 10-12 December, 2010. The author wish to thank the participants and Gian Luigi Mazzi for comments and useful discussion on the issues covered in the article.

References

Chow, G. & Lin, A. L. (1971), 'Best linear unbised interpolation, distribution and extrapolation of time series by related series', *The Review of Economics and Statistics* **53**(4), 372–375.

de Jong, P. (1991), 'The diffuse kalman filter', *The Annals of Statistics* **19**(2), 1073–83.

Durbin, J. & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

Fahrmeir, L. (1992), 'Posterior mode estimation by extended kalman filtering for multivariate dynamic generalised linear models', *Journal of the American Statistical Association* **97**, 501–509.

Fernández, F. & Harvey, A. (1990), 'Seemingly unrelated time series equations and a test for homogeneity', *Journal of Business and Economic Statistics* **8**, 71–81.

Fernández, P. E. B. (1981), 'A methodological note on the estimation of of time series', *The Review of Economics and Statistics* **63**, 471–478.

Frale, C., Marcellino, M., Mazzi, G. L. & Proietti, T. (2010*a*), 'Euromind: A monthly indicator of the euro area economic conditions', *Journal of the Royal Statistical Society (Series A - Statistics in Society)* . Forthcoming.

Frale, C., Marcellino, M., Mazzi, G. L. & Proietti, T. (2010*b*), 'Survey data as coicident or leading indicators', *Journal of Rorecasting* **29**, 109–131.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A. C. & Chung, C. (2000), 'Estimating the underlying change in unemployment in the uk', *Journal of the Royal Statistical Society A* **163**, 303–39.

Harvey, A. & Koopman, S. (1997), Multivariate structural time series models (with comments), *in* B. H. C. Heij, J.M. Shumacher & C. Praagman, eds, 'System Dynamics in Economic and Financial Models', Chichester: John Wiley and Sons Ltd., pp. 269–98.

Koopman, S. (1993), 'Disturbance smoother for state space models', *Biometrika* **80**, 117–26.

Koopman, S. J., Harvey, A. C., Doornik, J. A. & Shephard, N. (2009), *STAMP 8.2: Structural Time Series Analyser, Modeller and Predictor*, Timberlake Consultants Press.

Litterman, R. B. (1983), 'A random walk, markov model for the distribution of time series', *Journal of Business and Economic Statistics* **1**, 169–173.

Moauro, F. & Savio, G. (2005), 'Temporal disaggregation using multivariate structural time series models', *The Econometrics Journal* **8**, 214–234.

Proietti, T. (2006*a*), 'On the estimation of nonlinearly aggregated mixed models', *Journal of Computational and Graphical Statistics* **15**(1), 18–38.

Proietti, T. (2006*b*), 'Temporal disaggregation by state space methods: Dynamic regression methods revisited', *Econometrics Journal* **9**, 357–372.

Proietti, T. (2008), Estimation of common factors under cross-sectional and temporal aggregation constraints: Nowcasting monthly gdp and its main components. MPRA Paper No. 6860, http://mpra.ub.uni-meunchen.de/6860/.

Proietti, T. & Moauro, F. (2006), 'Dynamic factor analysis with non-linear temporal aggregation constraints', *Applied Statistics* **55**(2), 281–300.

Stock, J. H. & Watson, M. (1991), A probability model of the coincident economic indicators, *in* G. Moore & K. Lahiri, eds, 'Leading Economic Indicators', Cambridge University Press.

Table 1 Model estimates of temporal disaggregation regressions for total employment in the euro area. Models in the levels

	$\widehat{\mathcal{L}}$	$\widehat{\rho}$	$\widehat{c}$	$\widehat{g}$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
CL	-450.63	.9600	138733.64	118.52	-1.22	-
			(123.30)	(37.77)	(-15.07)	-
FE	-450.42	-	138234.62	113.45	-1.17	-
			(112.45)	(12.72)	(-13.21)	-
LT	-450.42	.0500	138207.12	113.38	-1.16	-
			(111.43)	(12.57)	(-13.07)	-
ADL(1,0)	-477.10	.8150	26385.49	21.58	-0.27	-
			(113.81)	(61.97)	(-16.12)	-
ADL(1,1)	-448.62	.9645	5286.43	3.98	-1.12	1.06
			(41.36)	(25.61)	(-12.21)	(11.24)
ADL(1,0)D	-470.70	.8725	-	-	9.72×10^{-4}	-
			-	-	(2.27)	-
ADL(1,1)D	-448.74	.0005	376.32	-0.37	-1.10	1.08
			(2.96)	(-1.75)	(-11.97)	(11.46)

Table 2 Model estimates of temporal disaggregation regressions for total employment in the euro area. Models in the logarithms

	$\widehat{\mathcal{L}}$	$\widehat{\rho}$	$\widehat{c}$	$\widehat{g}$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
CL	-452.33	.9795	0.26	1.8×10^{-5}	-0.11	-
			(162.14)	(25.07)	(-13.17)	-
LT	-451.76	.1855	6.95×10^{-4}	-	-0.11	-
			(11.99)	-	(-11.96)	-
ADL(1,0)	-480.59	.9955	0.13	-	-7.74×10^{-3}	-
			(10.87)	-	(-6.32)	-
ADL(1,1)	-450.30	.9735	0.36	2.2×10^{-5}	-0.10	0.10
			(42.25)	(20.71)	(-11.88)	(11.08)
ADL(1,0)D	-469.33	.8710	-	-	1×10^{-5}	-
			-	-	(2.37)	-
ADL(1,1)D	-451.95	.0300	-	-	-0.10	0.10
			-	-	(-12.45)	(12.46)

Table 3. Estimated standard deviations of disturbances. Simultaneous SUTSE model in logs

		Germany	Belgium	Euro area	Spain	France	Italy	The Netherlands
				QNA Employment $(\times 10^{-3})$				
AGR	level	2.0508	4.9246	3.7427	7.5077	1.0120	18.4810	3.5334
	irregular	14.4750	7.7023	9.5137	17.3530	0.9029	11.7420	20.5210
IND	level	2.2303	1.6770	1.6246	5.4994	1.7886	2.2930	2.2425
	irregular	2.0983	1.8722	1.6480	4.8281	1.4233	5.2712	3.3440
COS	level	2.3899	4.2063	3.3400	10.0060	3.2333	7.5036	3.8258
	irregular	2.2413	15.3860	3.7608	6.7306	2.1688	13.3280	6.6848
TTC	level	1.2532	1.4905	1.3935	3.5970	1.3105	3.4991	2.2352
	irregular	1.6484	2.9010	1.9715	4.0996	1.0455	12.3990	2.7869
FBS	level	2.7729	2.5354	2.2272	4.8225	2.6110	5.2697	4.7254
	irregular	2.8522	2.3884	2.2267	6.8751	2.6232	6.1178	4.9540
OTS	level	0.8011	1.1094	0.7664	2.5579	0.5523	3.0217	1.4548
	irregular	0.7730	2.7549	4.6074	13.1660	0.4235	24.1210	1.8611
				Monthly unemployment $(\times 10^{-3})$				
Total	level	1825.0	888.5	756.5	1513.7	724.6	2030.9	2531.7
	irregular	11.4190	5.0756	3.8357	7.8146	3.8271	22.3170	17.7320

Table 4. Level component cross correlations with euro area total employment. Simultaneous SUTSE model in logarithms

	Belgium	Germany	Euro area	Spain	France	Italy	The Netherlands
			QNA Employment				
AGR	0.055	0.126	0.579	0.324	0.411	0.542	0.123
IND	0.282	0.328	0.677	0.507	0.511	0.461	0.270
COS	0.269	0.146	0.739	0.729	0.416	0.187	0.369
TTC	0.110	0.284	0.597	0.519	0.427	0.254	0.279
FBS	0.403	0.528	0.696	0.620	0.608	0.300	0.294
OTS	0.019	-0.166	0.103	0.343	0.134	0.000	-0.004
			Monthly unemployment				
Total	-0.146	-0.289	-0.738	-0.516	-0.650	-0.420	-0.340

Table 5. Root mean squared revision errors in the estimation of quarterly growth rates of employment in the euro area. Regression models in logarithms.

	AGR	IND	COS	TTC	FBS	OS	Total
CL indirect	.7413	.5306	1.1103	.4935	.7064	**.3885***	.2879
direct							.2760
LT indirect	.8528	.4721	1.0568	.5026	.7309	.4188	.2800
direct							.2635
ADL(1,0) indirect	**.7390***	.5389	1.1598	**.4792**	.7024	.3928	.2785
direct							.2786
ADL(1,1) indirect	.7613	.5117	1.1042	.4834	.7003	.3921	.2797
direct							.2677
ADL(1,0)D indirect	.8162	.4980	1.0829	.5117	.7356	.4172	.2710
direct							.2722
ADL(1,1)D indirect	.8607	**.4691**	**1.0448***	.4924	**.6877**	.4162	.2690
direct							**.2560**

Table 6. Root mean squared revision errors in the estimation of quarterly growth rates of employment in the euro area. SUTSE models in logarithms.

	AGR	IND	COS	TTC	FBS	OS	Total
Sectoral approach with MS data							
SUTSE indirect	**.7870**	.4488	1.1193	**.4691***	.6764	.4129	.2772
direct							.2639
Contemporaneous modelling approach							
without MS data	.8153	.4648	1.0669	.5057	.6881	**.3940**	.2841
with MS data	.9025	**.4101***	1.0578	.4752	**.6506***	.4105	**.2553***

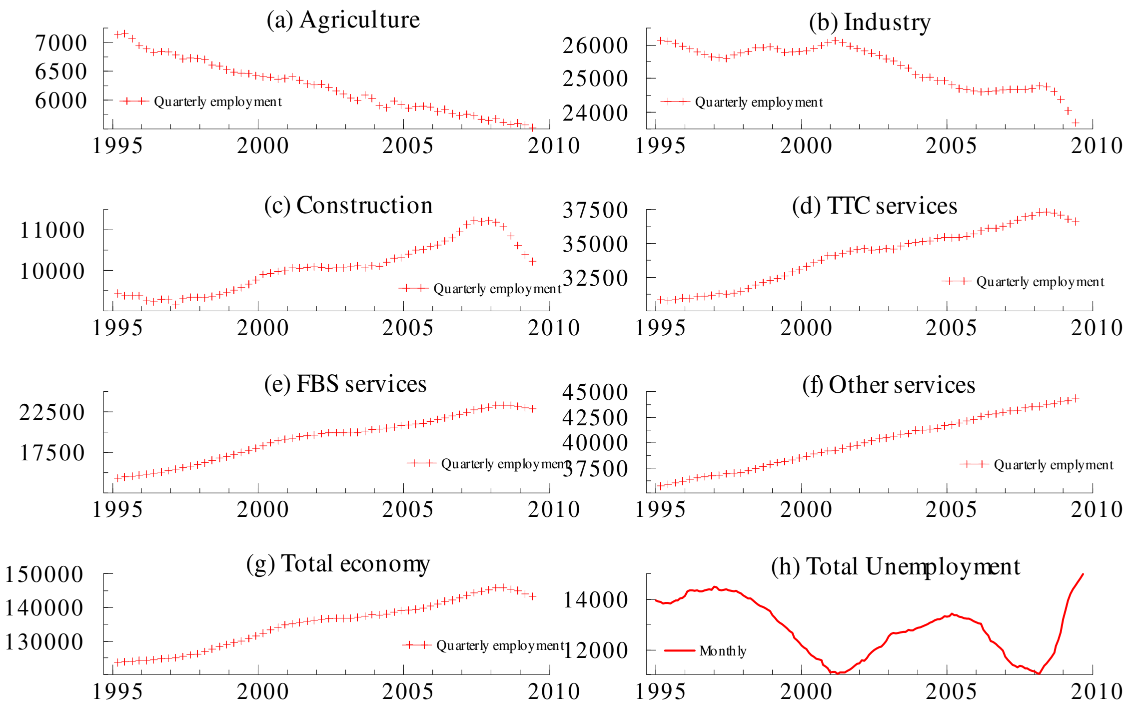

Figure 1: Quarterly employment and monthly unemployment in the euro area

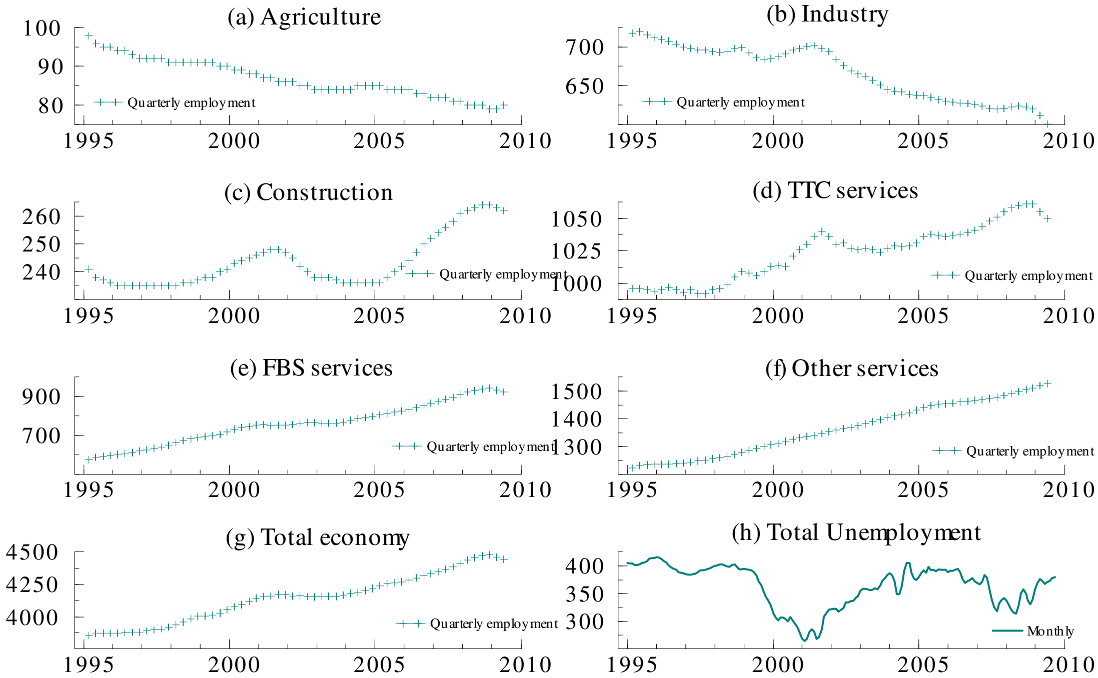

Figure 2: Quarterly employment and monthly unemployment in Belgium

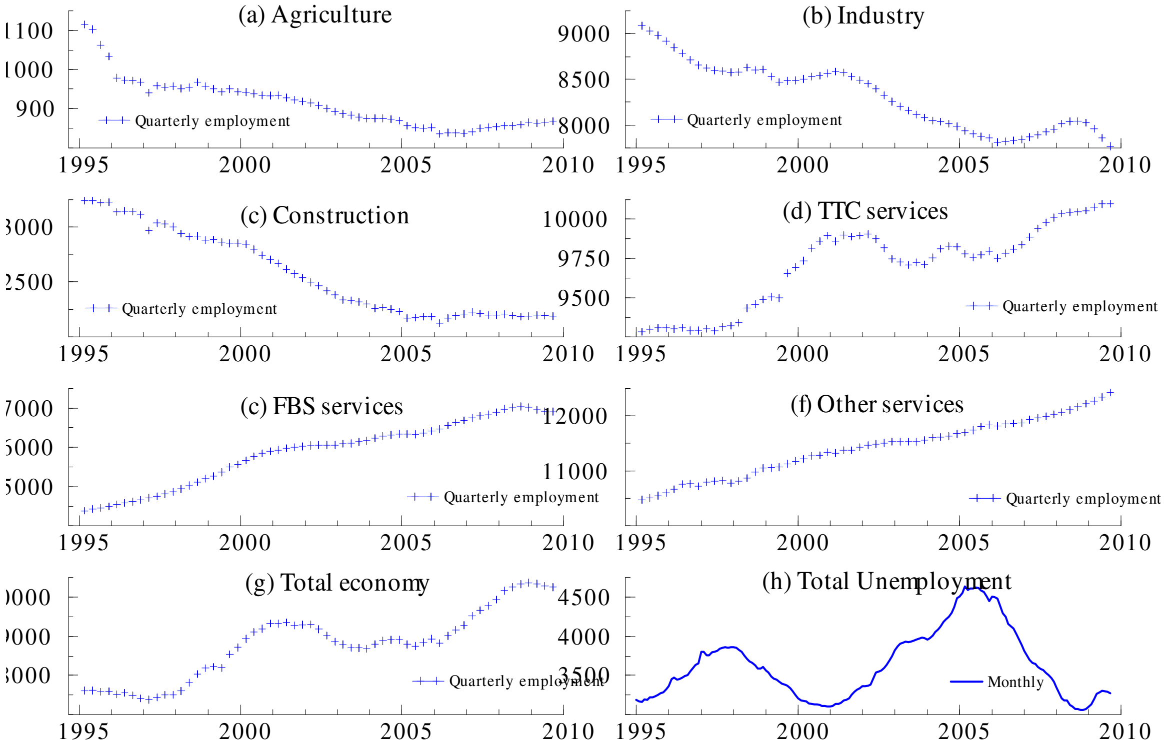

Figure 3: Quarterly employment and monthly unemployment in Germany

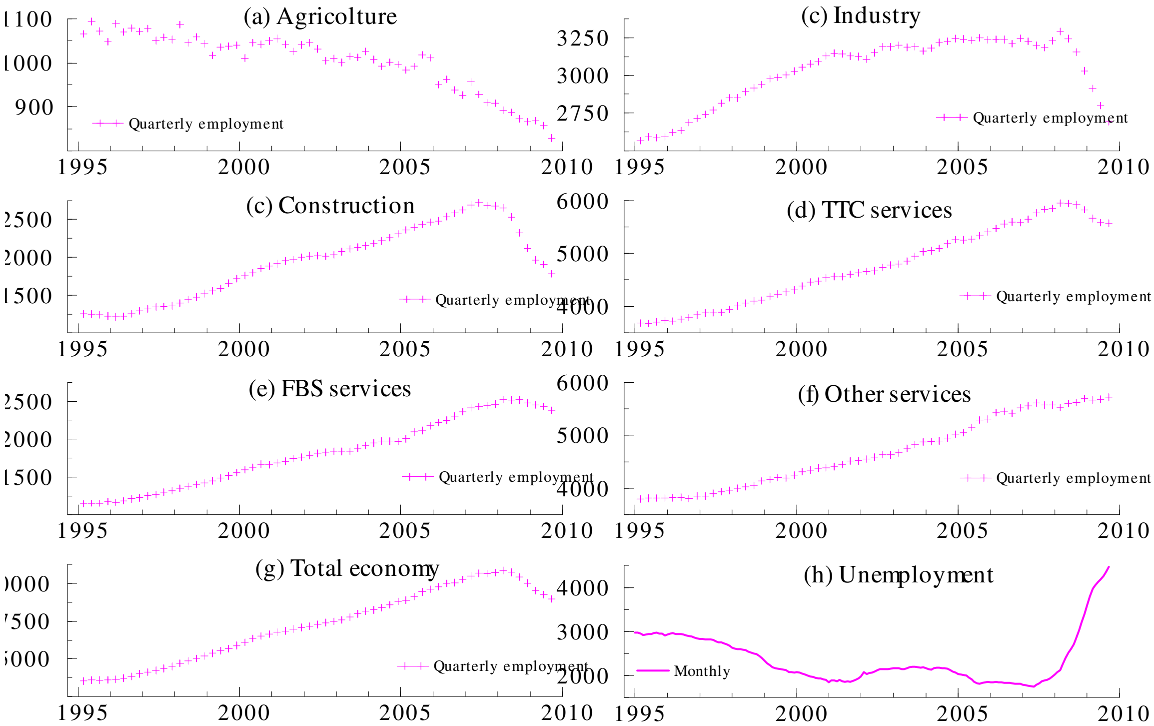

Figure 4: Quarterly employment and monthly unemployment in Spain

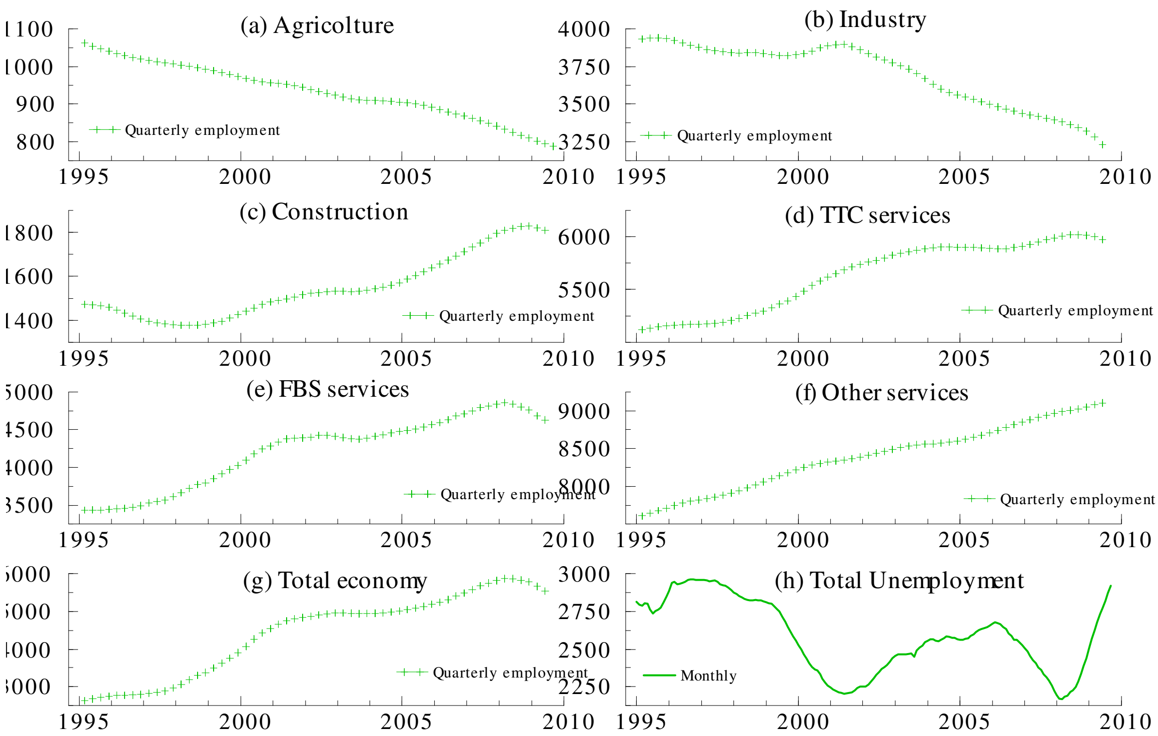

Figure 5: Quarterly employment and monthly unemployment in France

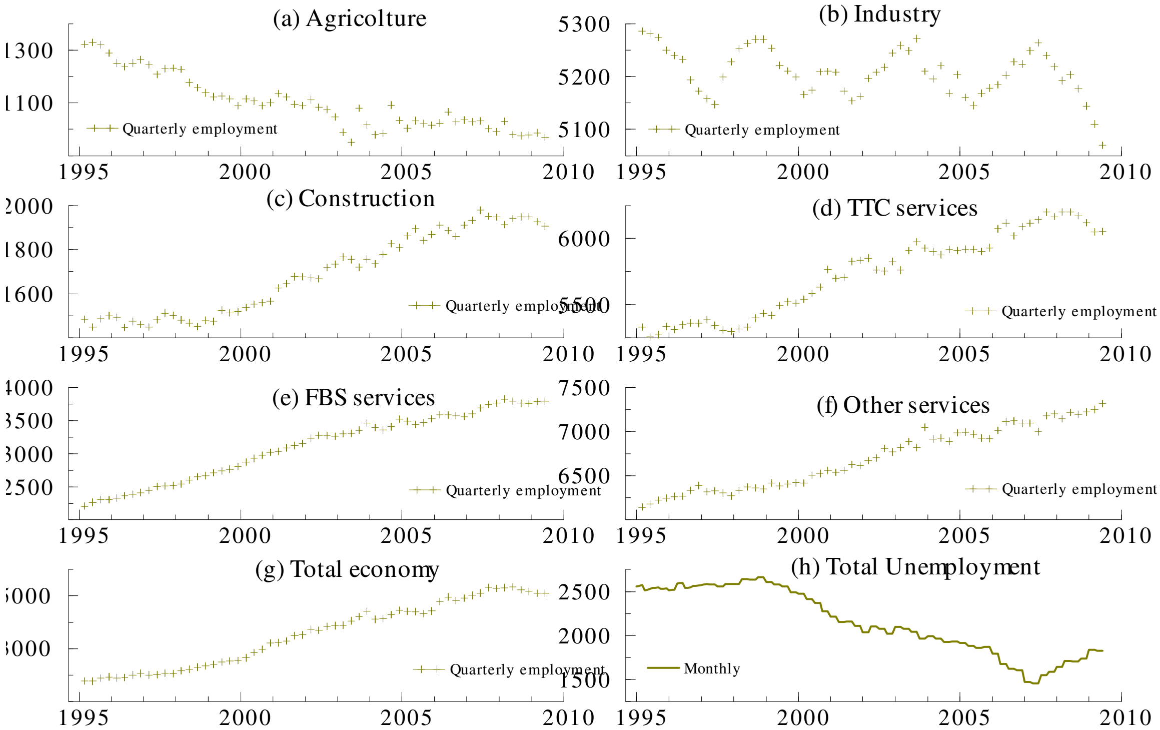

Figure 6: Quarterly employment and monthly unemployment in Italy

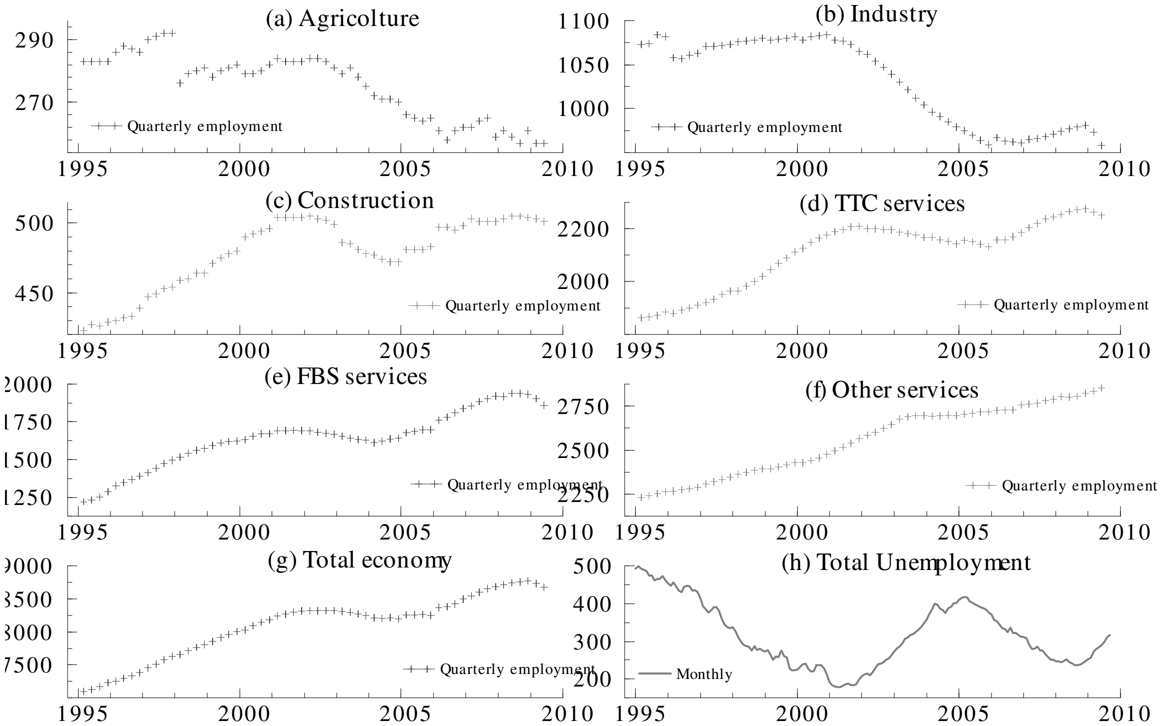

Figure 7: Quarterly employment and monthly unemployment in the Netherland

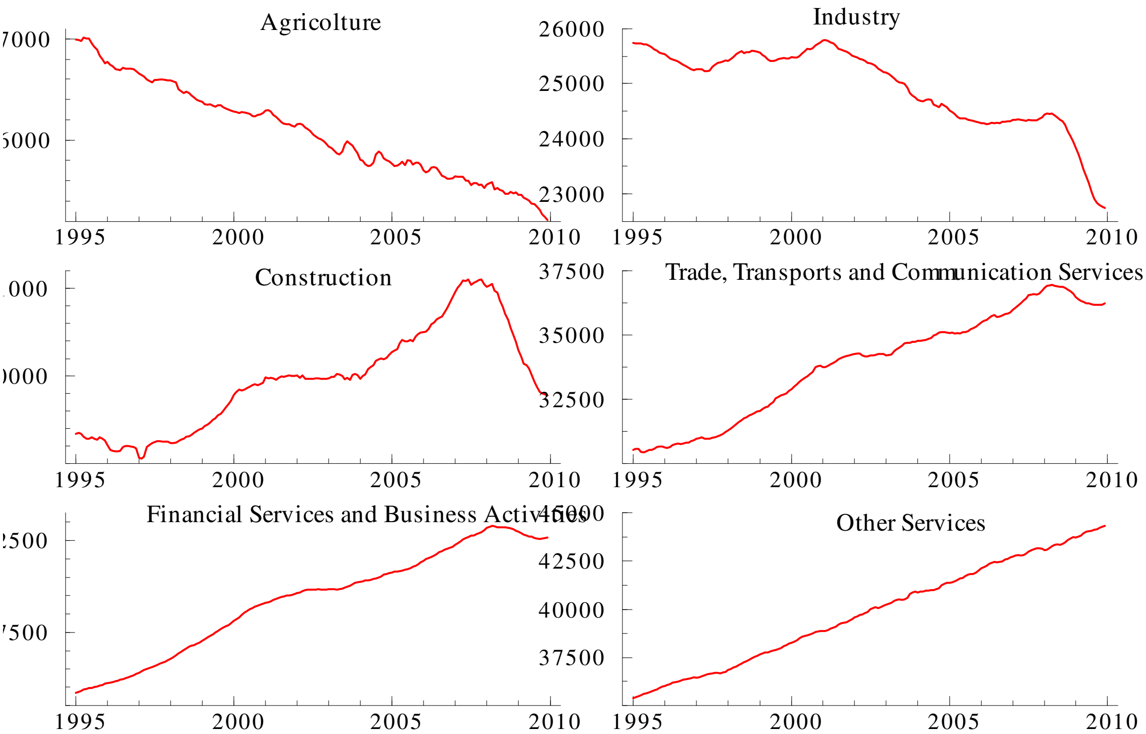

Figure 8: Monthly estimates of employment by sector in the euro area

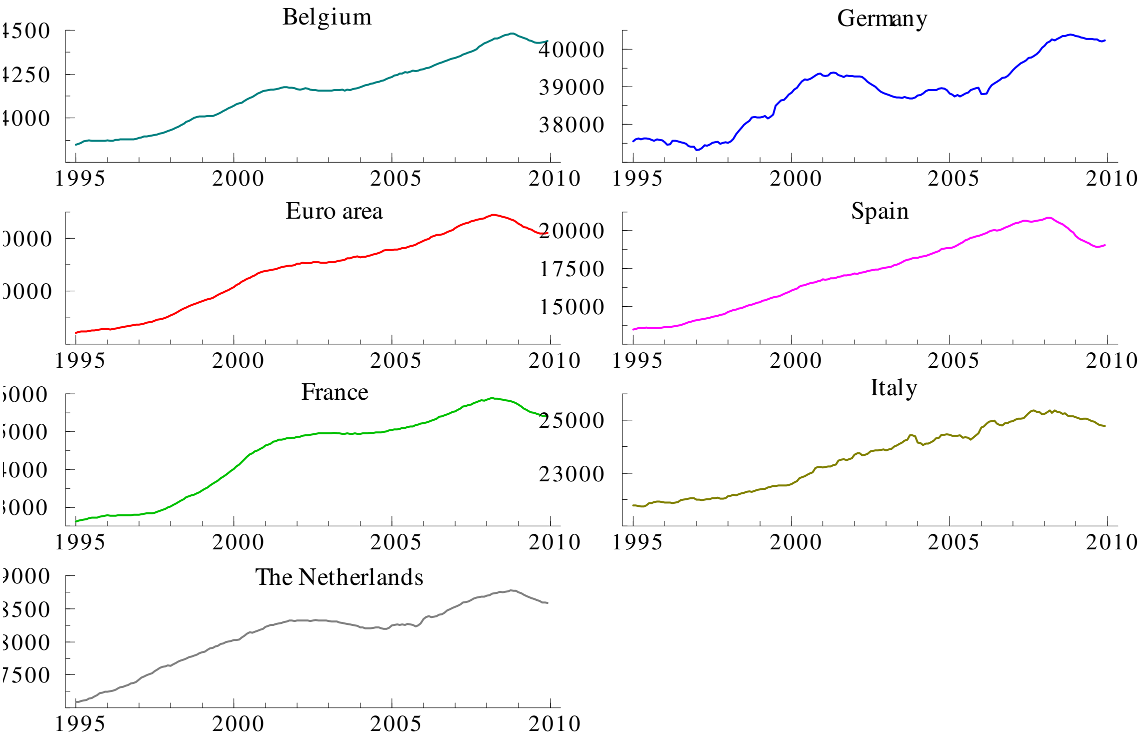

Figure 9: Monthly estimates of total employment in the euro area and Member States

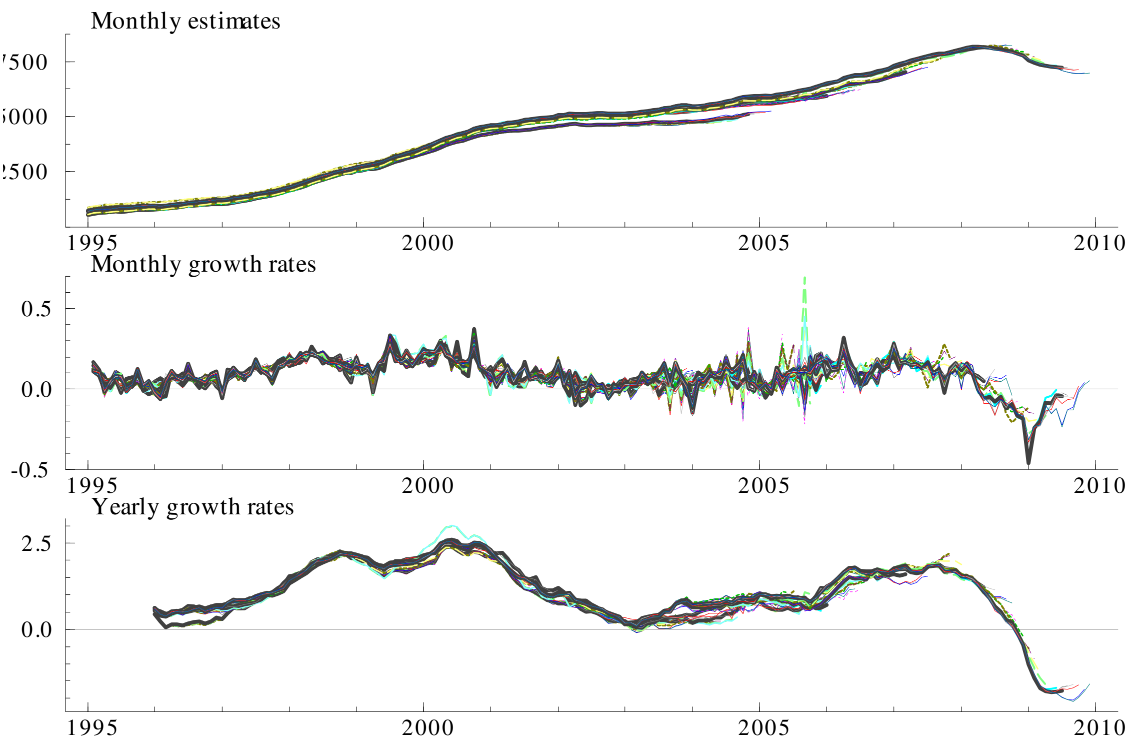

Figure 10: Monthly estimates of employment for the contemporaneaus SUTSE modelling approach - 73 vintages February 2004 - January 2010