Firms' Dynamics and Business Cycle: New Disaggregated Data*

LORENZA ROSSI[1] AND EMILIO ZANETTI CHINI[†1]

[1] *Department in Economics and Management, University of Pavia*

May 16, 2016

Abstract

We provide stylized facts on firms dynamics by disaggregating U.S. yearly data from 1977 to 2013. To this aim, we use a new unobserved component-based method, encompassing several classical regression-based techniques currently in use. The new time series of Entry and Exit of firms at establishment level are feasible proxies of Business Cycle. Exit is a leading and countercyclical indicator, while Entry is lagging and procyclical. The resulting SVAR analysis supports the recent theoretical findings of the literature on firms dynamics.

Keywords: C10, C13, E3, E32, E37.

JEL: Entry and Exit, State Space, Business Cycle, Disaggregation, SVAR.

*This work is supported by the EU 7th framework collaborative project "Integrated Macro-Financial Modelling for Robust Policy Design (MACFINROBODS)" grant no. 612796. The Authors are in debt with Tommaso Proietti for having shared his "TDS" Ox-package and his suggestions; obviously, any error or imprecision in the calculations is our responsibility. They also thanks Barbara Annicchiarico, Cecilia Susanna Jona-Lasinio, Jean Palate and all participants to ERCIM-CFE2015 in London and 2nd MACFINROBODS in Bruxelles.

†Corresponding Author: Via San Felice 5, 27100 Pavia. E-mail: ezc@eco.unipv.it .

1 Introduction

It is an established result in Macroeconomics that the entry and exit of firms play an important role for business cycle analysis. This topic is becoming particularly appealing in light on the recent theoretical advances in Dynamics Stochastic General Equilibrium (DSGE) models.[1] Hence, a full understanding of the dynamics underlining these variables is necessary to model - and possibly forecast - the Business Cycle. Anyway, the descriptive ability of macroeconomic time series data is still unclear and, more important, their time-span coverage heavy limited. This scarcity of macro data complicates the econometric analysis of the firm dynamics and constitutes the motivation of this paper. In this respect, our contribution is twofold. First, we improve the availability of the US data on firms' dynamics by disaggregating the only official yearly long-span dataset available for the U.S. economy from the US Census Bureau. Second, we provide a business cycle analysis of the new disaggregated series.

In the first part of this paper, we consider two different types of temporal disaggregation techniques: i) the one proposed by Chow and Lin (1971), which is the standard for many Statistical Institutes to estimate their economic indicators; and ii) two models based on unobserved components methods (UCM), originally proposed by Proietti (2006). The first model is simply based on an OLS estimation of an AR(1) process. On the other hand, the two models based on UCM are characterized by a unifying state space representation and capable to model a more rich dynamics. Their estimation relies on an Augmented Kalman Filter (AKF). Thanks to this, the UCM allows generality and flexibility besides maintaining statistical robustness.

We find that the resulting series disaggregated by UCM are considerably more accurate and credible of the ones resulting by applying the *naive* Chow-Lin model. Furthermore, the selecting procedure of the two models using the UCM suggests that series of establishments exit is more accurately fitted by an UCM representa-

[1] see for example the seminal paper of Bilbiie et al (2012) among others

tion of an AR(1) model, while the series of establishment entry is better represented by an UC representation of an ADL(1) model. Both the models estimated via UCM are univariate, and this means that their underlining regression models rely on a single regressor, here represented by the industrial production index. The use of a single regressor might be sub-optimal in a disaggregation exercise and constitutes the main limitation of this class of univariate techniques. A first-attempt to circumvent this problem, here, is to re-apply extensively the new (univariate) UCM on all the indicators of a new macroeconomic dataset. The disaggregated variables of entry and exit are then computed as a simple average of the resulting univariate estimates. In this way, we make use of all possible informations on the US economy when extracting the quarterly estimates. This lead to new entry and exit series, qualitatively similar to the ones derived from the single indicator.

The second part of this paper deals with the business cycle analysis. Namely, we extract the trend and cycle components via standard Hodrick-Prescott (HP) and Baxter-King (BK) filters. Once the cycle component has been extracted, we classify the disaggregated series as leading or lagging indicators of the business cycle by looking at the maximum absolute value of cross-correlations between the cycle of disaggregated entry and exit and that of the real GDP. According to our results, entry is a lag and pro-cyclical indicator, while, on the contrary, exit is countercyclical leading.

Finally, in the last part of this paper, we run a Structural Vector AutoRegression (SVAR) analysis and shows the responses of the series of entry, exit and of real GDP to a shock on the total factor productivity (TFP). This shock is identified simply using the short run restriction that the TFP is the most exogenous variable. The impulse response functions (IRF) of the resulting SVAR model seems compatible with recent macroeconomic literature on endogenous firms dynamics; see, for example, Bilbiie *et al.* (2012); Rossi (2015); Lewis (2009); Etro and Colciago (2010); Colciago and Rossi (2015) among others. Indeed, we find that a TFP shock is followed with a negative and persistent response of exit and a positive and persistent response of entry, together with a positive response of the real GDP. Importantly, the comparison with the IRFs obtained estimating the same VAR using the BLS quarterly series suggests that our disaggregated series are good proxies of firms dynamics. Our series offer the advantage of considering a longer period which is almost twice the size of the sample of the quarterly series.

This paper is related to the recent DSGE literature on firms dynamics. The theoretical contributions on firms' dynamics are mainly focused on firms's entry. In their seminal article, Bilbiie *et al.* (2012) - henceforth, BGM - introduce a DSGE model with endogenous firms' entry, according to which the sluggish response of the number of producers, due to the sunk entry costs, generates a new, potentially important endogenous propagation mechanism for real business cycle models; see also Bergin and Corsetti (2008); Jaimovich and Floetotto (2008); Etro and Colciago (2010); Colciago and Rossi (2012); Lewis and Poilly (2012); Siemer (2014); Bergin *et al.* (2014); Casares and Poutinau (2014); La Croce and Rossi (2015). These papers consider an exogenous and constant exit probability of firms from the market. Thus, they cannot disentangle the role of firms exit with respect to firm's entry, omitting, in such a way, an important feature of business cycle. The role of firms exit has been more recently considered by Rossi (2015) and Hamano and Zanetti (2015). Both these theoretical papers show that firms' exit represents an even stronger propagation mechanism of the business cycle than that of firms' entry. As previously mentioned, these efforts to give a theoretical explanation of business dynamics face with severe data limitations. Fortunately, our disaggregated series conveys a reliable picture of establishment dynamics and represent a step ahead for evaluating theoretical models studying the impact of firms' dynamics on business cycle.

The rest of this paper is organized as follows. Section 2 deals with the problem of data availability on firms/establishment dynamics. Section 3 describes the methodology. Then, Section 4 applies the disaggregating techniques on US data on establishments entry and exit and investigates the business cycle properties of the disaggregated series. Section 6 discusses the relevance of their business cycle movements for Macroeconomics via a SVAR analysis. Finally, Section 7 concludes. Technical details of the methods used are left in the Appendix.

2 Data on Firms/Establishment Dynamics

The only official long-span dataset on firms' dynamics is the Business Dynamics Statistics (BDS), published by Census Bureau Research Data Centers. It gives information about total number of firms, establishments and workers, establishments opened, establishments closed, job creation and job destruction and other measures at yearly frequency from 1977 up to 2013. In turn, BDS data summarize the confidential data of Longitudinal Business Databases, a census of business establishments and firms in the US covering all industries and all US[2]. In particular, according to BDS an establishment opening or entrant is an establishment with positive employment in the current year and zero employment in the prior year. An establishment closing or exit is an establishment with zero employment in the current year and positive employment in the prior year. The vast majority of establishment openings (closings) is constituted by true greenfield entrants (exits). However, a small number of them are temporarily shutdown (i.e., have a year with zero employment[3]); these are excluded from the counts of establishment openings and closings. In the course of this paper we will refer to these series with the "ENTRY" and "EXIT" labels.

For higher frequency data, two are the main sources available for applied analysts:

1. BUSINESS EMPLOYMENT DYNAMICS from Bureau of Labor Statistics (BLS, henceforth). This source provides a quarterly census of the labor force in private establishments from 1992:Q3 and measures the net change in em-

[2] For more details, see the, Census web page, where data are available.

[3] "Zero employment" means that an employment level of zero was reported, whereas "no employment" means that there were not any employment numbers reported at all.

ployment at the establishment level. According to the BLS definition, a net increase (decrease) in employment comes from opening (closing) establishments. Openings are either establishments with positive third month employment for the first time in the current quarter, with no links to the prior quarter, or with positive third month employment in the current quarter following zero employment in the previous quarter. According to the BLS definition openings include both new startups (births) and re-openings of the existing seasonal establishments that reported zero employment in the previous quarter. Closings are either establishments with positive third month employment in the previous quarter, with no positive employment reported in the current quarter, or with positive third month employment in the previous quarter followed by zero employment in the current quarter. Closings include establishments that go out of business permanently (deaths), as well as seasonal businesses that shut down temporarily. We will refer to these series with the "OPENINGS" and "CLOSINGS" labels.

Alternatively, it is also possible to use the two series of establishment births and deaths provided by the Business Employment Database (BED) of BLS for total private sector, available from the same source from 1993:Q2. In particular, for the purpose of BED statistics, births are defined as establishments that appear in the longitudinal database for the first time with positive employment in the third month of a quarter, or showed four consecutive quarters of zero employment in the third month followed by a quarter in which it shows positive employment in the third month. Births are a subset of openings since they not include re-openings of seasonal businesses. Similarly, deaths are defined as establishments that either drop out of the longitudinal database or an establishment that had positive employment in the third month of a given quarter followed by four consecutive quarters of showing zero employment in the third month. Differently from closings, deaths do not include temporary shutdowns. This means that they are also a subset of closings. For this reason, their dynamics is much less volatile than that of openings and closings. We will refer to these series with the "BIRTHS" and "DEATHS" labels. As we will show in this paper, this will imply important consequences for the business cycle analysis of these series.

2. ECONOMAGIC. It provides a monthly series on the number of new business incorporations from 1959:M1 to 1996:M9.

Despite the similar nature of the data, and their temporal contiguity, these two series measure different objects. Incorporations concern firm's level of aggregation, while, on the other hand, establishments are often partitions of the firm. Consequently, the observations have different order of magnitude: if one aggregates the Economagic' s monthly series, the resulting values are, approximately, a half of the observed BLS ones. Hence, no interpolation is possible between the BLS and Economagic series. This implies that the above-mentioned macroeconomic literature is forced to stay to the evidence of short samples, or to make theoretical assumptions on the law of motion of firms' entry without any empirical counterfactual. Perhaps more astonishing, the exact definition of incorporation used is not available, despite the fact that such series is the basis of a recent strand of literature.[4]
Thus, firms' dynamics is not observed and is measured by a proxy - the employment level - quite imperfect.

3 Temporal Disaggregation

Temporal disaggregation, from an operational point of view, can be seen as a specular version of the problem of temporal aggregation of macroeconomic variables. In a disaggregation exercise, measurements of stock/flow variables are available only

[4] See Bergin and Corsetti (2008); Lewis and Poilly (2012); Bergin *et al.* (2014); Lewis (2013) among others.

over s consecutive periods, where $s = 4$ if moving from yearly to quarterly, 12 from yearly to monthly and so on. In our case, the annual total of a macroeconomic (flow) variable (i.e. ENTRY and EXIT) has to be redistributed across the quarters using related series that are available at higher frequency (indicators).

More in details, the problem of temporal disaggregation is solved via *interpolation* or *distribution*. The former consists of the estimation of the missing values of a *stock variable* at points in time that have been systematically skipped by the observation process. The latter arises when *flow variables* are in the form of linear aggregates, as for the case of observations available only as totals or as averages over s consecutive periods. Since establishments ENTRY and EXIT are flow variables, temporal distribution represents exactly the solution for our disaggregation problem. This Section presents the mostly used disaggregation methods. First, we briefly introduce the Chow and Lin (1971) method. This relies on a simple OLS regression of an AR(1) process, where the regressors are observed indicators, whose distribution function is assumed to be known. For this reason we label this method as the Chow-Lin Naive Method (CL-NM). We then discuss the two models based on UCM. The first one is a UC representation of the Chow-Lin regression model (CL-UCM) and second one is a more general autoregressive distributed lag (ADL-UCM) model. Both of them are estimated by AKF[5].

The use of the univariate methods implies that the resulting estimates can be highly affected by the choice of the indicator. To overcome this problem, in the final part of this Section we propose to apply the univariate UCM to all the 134 indicators of MD-FRED dataset, a new macroeconomic database of 134 monthly indicators of U.S. economic activity which has been recently published by Federal Reserve Bank of St. Louis and extensively described in McCraken and Ng (2015)[6]. This dataset will be used to repeat the disaggregation via UCM - for all the indicators. The

[5]All the models considered rely on a single regressor, or indicator variable, represented by the series of Industrial Production Index, downloaded from FRED of St. Louis.

[6]The FRED-MD is at monthly frequency. Since in our application the focus in on quarterly series, we aggregated the original data available.

disaggregated variable will be then computed as a simple average of the resulting 134 univariate UCM estimates. We label this method as Combination Method.

3.1 The CL-NM Method

The first technique to deal with temporal disaggregation of flow variables has been introduced by Chow and Lin (1971) (CL-NM). In the case of univariate estimation, this is a simple linear OLS regression of the observed process y_t on the vector of indicator variable x_t multiplied by the disaggregation matrix C_d, i.e.

$$C_d y_t = C_d x_t \beta + C_d u_t \tag{1}$$

where $u_t \sim N\left(0, C_d V C_d'\right)$ is a vector of autocorrelated errors of order 1, i.e.

$$u_t = \phi u_{t-1} + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma^2) \tag{2}$$

with $|\phi| < 1$, and x_t is assumed to be: i) exogenous, ii) free of measurement errors iii) cointegrated. Notice that, for the disaggregation from yearly to quarterly observations the matrix C_d is of the form

$$C_d = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

The model is estimated using the standard OLS method. As stressed by Harvey (1989) assumptions i) to iii) represent strong limitations of this regression based method. To overcome these limitations Proietti (2006) adapts the UCM, originally introduced by Harvey (1989), to the disaggregation techniques. The next sub-Section discusses these techniques.

3.2 The Unobserved Component Method

Let y_t represents a $[T \times 1]$ vector of unobserved elements dating from $1, \ldots, t, \ldots, T$. Then the general model for temporal disaggregation relies on the following state space representation:

$$\begin{cases} y_t = z'\alpha_t + x_t'\beta, \\ \alpha_t = \alpha_{t-1} + W_t'\beta + H\epsilon_t \\ \alpha_1 = a_1 + W_1\beta + H\epsilon_1 \end{cases} \tag{3}$$

The first equation is named *measurement equation*, while the second one is the *transition equation*. The vectors x_t and the matrices W_t contain exogenous regressors that enter respectively the measurement and the transition equations and zero elements corresponding to effects that are absent from one or the other equations. They are usually termed *indicators* in the literature. The initial state vector, α_1, is expressed as a function of fixed and known effects (a_1), random stationary effects ($H_1\epsilon_1$, where the notation stresses that H_1 may differ from H), and regression effects, $W_1\beta$.

For what follows the following assumptions are invoked:

Assumption 1. $Var(\beta) \to 0$.

Assumption 2. $Var(\beta)^{-1} \to 0$.

Assumption 1 means β is fixed, but unknown. It holds if it is deemed that the transition process governing the states has started at time $t = 1$. Assumption 2 implies that β has an improper distribution with mean 0 and arbitrarily large variance matrix. This holds if the process has started in the indefinite past.

Both of these assumptions are just for convenience and can easy be weaken. Under this general framework, the Chow-Lin can be assumed as particular case, albeit differently from the CL_NM it relies on unobserved component techniques.

Case 1 (CL-UCM model)**.** When α_t is a scalar, $z = 1$, $T = \phi$ are $H = 1$, the system

(3) degenerates into a linear regression model with AR(1) errors:

$$y_t = z'\alpha_t + x_t'\beta, \quad \alpha_t = \phi\alpha_{t-1} + \epsilon_t \quad, \epsilon_t \sim NID(0, \sigma^2) \tag{4}$$

with $\phi < 1$, $\alpha_1 \sim N(0, \sigma^2(1-\phi^2))$.

Remark 1. The CL-NM model assumes full cointegration of non-stationary elements eventually present in x_t. In this case the deterministic component is handled via inclusion of regressors as, e.g., $x_t = 1, t, x_{3t}, \ldots, x_{kt}]'$ and re-arrangement of the process as $y_t = \mu_t + \gamma t + \sum_j \beta_j x_j + \alpha_j$ with μ and γ being the first two elements of β.

Case 2 (ADL-UCM model)**.** If $z' = 1$, $T = \phi$, $H = 1$, $W = [1, t, x_t', x_{t-1}']$, $\beta = [m, g, \beta_0', \beta_1']'$, the system (3) degenerates into an ADL model:

$$y_t = \phi y_{t-1} + m + gt + x'\beta_0 + x_{t-1}'\beta_1 + \epsilon_t, \tag{5}$$

which initial conditions are, under stationarity assumption: $\alpha = 0$, $W_1 = \frac{1}{1-\phi}[1, \frac{1-2\phi}{1-\phi}, x_1', x_1']$, $H_1 = \frac{1}{\sqrt{1-\phi^2}}$. Still, changes are needed if assuming non-stationarity. See Proietti (2006).

Notice that the more general ADL-UCM can nest also the CL-UCM. Indeed:

Remark 2. The ADL-UCM collapses to the CL-UCM if and only if:

$$\beta_1 = \phi\beta_0. \tag{6}$$

Thus, (5) can be rewritten as:

$$y_t = x_t'\beta_0 + \alpha_t, \quad \alpha_t = \phi\alpha_{t-1} + \epsilon_t \tag{7}$$

so that the ADL(1,1) model nests a CL model with stationary AR(1) errors. Similarly, when $\beta_1 = 0$, the model become an ADL(1,0).

3.2.1 State Space Representation

Temporal disaggregation arises as a modification of system (3). In order to understand this, notice that (low-frequency) data are a sum of s consecutive values, available at time $t = 1, 2, \ldots, [n/s]$, where $[\cdot]$ denote the integer part of the number n/s. Then the following holds:

Proposition 1. Let define the cumulator variable as:

$$y_t^c = \psi_t y_{t-1}^c + \psi_{t-1}, \quad \psi_t = \tag{8}$$

Then the State Space System

$$\begin{cases} y_t = z^{*'} \alpha_t^* \\ \alpha_t^* = T_t \alpha_{t-1}^* + W_t^* \beta + H^* \epsilon_t, \\ \alpha_1^* = a_1^* + W_1^* \beta + H_1^* \epsilon_1, \qquad \epsilon_t \sim NID(0, \sigma^2), \quad \beta \sim N(b, \sigma^2 V), \end{cases} \tag{9}$$

where:

$$z^* = \begin{bmatrix} 0' \\ 1 \end{bmatrix}, \; T_t^* = \begin{bmatrix} T & 0 \\ z'T & \psi \end{bmatrix}, \; W_t^* = \begin{bmatrix} W_t \\ z'W_t + x_t' \end{bmatrix}, \; H_t^* = \begin{bmatrix} H_t \\ z'H_t + x_t' \end{bmatrix}, \tag{10}$$

$$\alpha_1^* = \begin{bmatrix} a_1 \\ z'a_1 \end{bmatrix}, \; W_1^* = \begin{bmatrix} W_1 \\ z'W_1 + x_1' \end{bmatrix}, \; H_1^* = \begin{bmatrix} H_t \\ z'H_1 + x_1' \end{bmatrix} \tag{11}$$

converts the disaggregation into a problem of estimation of a latent component model with missing observation. This is the system which is estimated for disaggregation of ENTRY and EXIT series, as shown in next Section 4.

Proof. This occurs simply replacing $y_t = z\alpha_t$ in (8), substituting the transition equation and re-writing:

$$y_t^c = \psi_t y_{t-1}^c + \psi_{t-1} + z'T\alpha_{t-1} + (z'W_t + x_t\beta) + z'H\eta_t, \tag{12}$$

and finally, substituting this expression in the state vector, with $\alpha_t^* = [\alpha_t', y_t^c]'$ $\square$

3.2.2 Estimation

Estimation of system (9) is done by Maximum Likelihood. Depending on the presence of diffuse elements in matrix W or not, two estimators are possible:

Fixed β: the ML estimators of β and σ^2 are:

$$\hat{\beta} = -S_{n+1}^{-1}s_{n+1} \quad , Var(\hat{\beta}) = S_{n+1}^{-1}, \quad \hat{\sigma}^2 = \frac{q_{n+1} - s_{n+1}'S_{n+1}^{-1}s_{n+1}}{[n/s]}, \tag{13}$$

with profile Likelihood:

$$\mathcal{L}_{\mathcal{F}} = -0.5[d_{n+1} + [n/s](\ln\hat{\sigma}^2) + (\ln\hat{\sigma}^2 + \ln 2\pi + 1)]. \tag{14}$$

Diffuse β: $\hat{\beta}$ and $\hat{\sigma}^2$ unmodified and $\hat{\sigma}^2 = \frac{q_{n+1} - s_{n+1}'S_{n+1}^{-1}s_{n+1}}{[n/s]-k}$ and the diffuse profile:

$$\mathcal{L}_{\mathcal{D}} := -0.5[d_{n+1} + [n/s - k](\ln\hat{\sigma}^2) + (\ln\hat{\sigma}^2 + \ln 2\pi + 1) + \ln|S_{n+1}|] \tag{15}$$

Notice that both of them, the parameters β can be concentrated out of the likelihood function, whereas the diffuse case is accommodated by simple modification of the likelihood.

Remark 3. S, s, q are outcomes of the augmented Kalman Filter (KF). This algorithm, introduced by de Jong (1991), enables exact inferences in the presence of fixed and diffuse regression effects. In order to make it operational in our system (3), the usual KF equations are augmented by additional recursions which apply the same univariate KF to k series of zero values, with different regression effects in the state equation, provided by W_t. The outputs vectors and matrices of the augmented KF are the basis for the de Jong (1989) augmented Smoothing algorithm. This refers to the estimation of the state vector α_t and the disturbance vector u_t using information in the whole sample rather than just past data. Smoothing is an important feature because it is the basis for diagnostic checking for detecting and distinguishing between outliers and structural changes using auxiliary residuals. Appendix A provides mathematical details.

3.3 Combination Method

The UCM disaggregation previously described holds for a single time series. In terms of system (9) this implies we assume x_t is $[T \times 1]$ vector. This assumption is often hard to justify in practical applications, being the resulting disaggregated time series y_t potentially too much depending from the indicator variable.

In order to mitigate this problem, we consider $\mathbf{X_t} = [\mathbf{x_1}, \ldots, \mathbf{x_n}, \ldots, \mathbf{x_N}]$ a $[T \times N]$ matrix containing the N indicators of MD-FRED dataset. In order to make use of all possible information from these data, we simply run the UCM for disaggregation machinery to all the N elements of X_t. Consequently, this means to produce N differently disaggregated processes collected in $[T \times N]$ matrix $\hat{\mathbf{Y}}_\mathbf{t} = [\hat{\mathbf{y}}_\mathbf{1}, \ldots, \hat{\mathbf{y}}_\mathbf{n}, \ldots, \hat{\mathbf{y}}_\mathbf{N}]$, with y_n representing the $[T \times 1]$ vector of the process disaggregated via system (3) which in turn corresponds to the $n-$th indicator of X_t and "$\hat{}$" denoting the fact that the series is the result of an estimation..

Then the combination of all disaggregated processes contained in Y_t is a simple average:

$$\hat{y}_C = \frac{1}{N} \sum_{n=1}^{N} y_n, \tag{16}$$

The equation (16) represents, in our applications, the process labelled as "AVEntry" or "AVExit".

4 Results

4.1 Comparison between univariate disaggregations

In this Section we apply the univariate CL-NM and the UCM to the BDS series of ENTRY and EXIT. Figure 1 shows the disaggregated series for all the three models considered together with the NBER recessions (in grey). Notice that the simple CL-NM leads to completely different results at the end of the sample (for example in 2013 the annual average level of EXIT is 170,000 in the CL-NM against 188,000 for the series measured by the equivalent CL-UCM method). These findings are symptomatic of the strong assumption considered by the CL-NM technique, particularly for the assumption of the existence of perfect cointegration between independent variable and the regressor (i.e. the industrial production). As a proof of this fact, Table 1 report the Johansen (1991) test for the null hypothesis - which is strongly rejected - of perfect cointegration, that is of having the matrix $Z = [x : y]$ not full rank (that is less than 2 in our application). This closes our investigation on Naive methods, because the full cointegration between time series is required, see Remark 1 in Section 3.2.

Another difference between the Naive method and UCM clearly emerges when looking at the correlations, reported in Table 2, between the quarterly BLS data of establishments OPENING (BIRTHS) and ENTRY series, jointly with the correlations between CLOSING (DEATHS) and EXIT. Notice that both the correlation between CL-NM ENTRY and OPENING (BIRTHS) is lower than the corresponding values obtained when their equivalent are computed using the UCM series. The same qualitative result holds for the correlations between CLOSING (DEATHS) and EXIT series. The latter presents correlations with the BLS series stronger than the ones obtained between OPENING (BIRTHS) and ENTRY series.

Further, notice that series estimated via CL-UCM specification are nicely smooth. Interestingly, such smoothness is lower if ADL specification is selected. This is immediately visible during NBER recession periods. Let consider the example, ENTRY during the recession phase of 1990: the series measured via CL-UCM passes from 178,000 in 1990:Q4 to 172,000 in 1991:Q2, while ADL-UCM series ranges from 181,500 to 167,000 (that is, the recession phase is more than the double in absolute value). This noisy behavior of ADL specification characterizes all the span of the series.

The finding of a non negligible difference of the two specifications of the UCM leads to a problem of selection of the regression model underlining disaggregation: what specification of UCM has to be chosen? We answer to this question by looking at the LR test for the null hypothesis that $\phi = 0$ in (6) versus the alternative hypothesis of $\phi \neq 0$ in the same equation, corresponding to the ADL(1,1) model[7]. Table 3 reports the parameter estimates of models (4) and (5) shows the one degree-of-freedom LR test. ENTRY has a LR statistics of 4.3, an this leads to its rejection. In other words, the model chosen for the ENTRY series is the ADL–UCM. On the contrary, for EXIT, the LR statistics of 1.5 is not able to reject the CL specification.

4.2 Combined Disaggregated Series

We use the result of LR-test as indication in order to select the specification to use when combining all the disaggregated series estimated with the same UCM from the 134 indicator variables in FRED-MD dataset. This lead to new ENTRY and EXIT series, qualitatively similar to the ones derived from a single indicator previously analyzed. These new series are labeled "AVEntry" ("AVExit") to underline the fact that they are an average of many single series, are then used for Business Cycle analysis. The next section extract the business cycle of the disaggregated series, both the ones obtained using the univariate method and the combined method. As Figure 3 seems to suggest, there are no sensible difference between the disaggregated series obtained using the univariate methods and that using the combined method.

[7]Also, all the diffuse effects x_t are reported, jointly with their standard deviation and t-statistics.

To further investigate the differences between the two types of disaggregated series and to analyze their short run dynamics, the next section extracts the business cycle of these series, analyzes their behavior with respect to the cycle of the RGDP and compares the dynamics of the disaggregated series with the one of their BLS equivalent.

5 Business Cycle Analysis

We analyze the business cycle dynamics of the disaggregated series by extracting the trend and the cycle components via standard filters. In particular, we consider the HP filter with penalization parameter $\lambda = 1600$ and BK filter with parameters $\lambda_0 = 1.5$, $\lambda_1 = 8$ and $K = 12$ (that is, $3 \times s$). Figure 4, panel (a) shows the resulting filtered series of Real GDP. In this case, the BK filter conveys a considerably smooth cycle with respect to the HP filter. On the contrary, BDS ENTRY and EXIT series obtained via both univariate and Combination UCM are almost coincident; only a weak difference between univariate and Combination methods is visible in terms of trends.

The same analysis is also performed for BLS variables. Figure 6 compares the two alternative measures of Entry and Exit, namely OPENINGS and BIRTHS - panel (a) - and CLOSINGS and DEATHS - panel (b). The HP-filtered OPENINGS series shows a wildly noisy trend and short cycle (very similar to the one of BIRTHS), while the CLOSINGS and DEATHS are more nicely smooth and different in their peaks and tougths; it is interesting to notice that the fall BIRTHS and the increase in DEATHS are almost than the double of their equivalent OPENINGS (CLOSINGS) in the 2007–2009 recession. This is due to the fact that OPENINGS include both new startups (births) and re-openings of the existing seasonal establishments that reported zero employment in the previous quarter. Similarly, CLOSINGS include also establishments that temporary shut-down. The series of BIRTHS and DEATHS instead do not suffer of this problem. For this reason, in what follows we compare the dynamics of the disaggregated ENTRY and EXIT (that by construction do not include infra-year reopening or temporary shut down) with the series of BIRTHS and DEATHS.

Figure 6 shows the business cycle of the disaggregated series of ENTRY and EXIT (using both the HP and the BK filter) and that of BIRTHS and DEATHS together the cycle of the RGDP, for the same sample period. The quarterly series of BIRTHS and DEATHS are more volatile and noisy than ENTRY and EXIT, and their cycle seems relatively shorter, even though overall they are qualitatively similar to the disaggregated series. Importantly, the same ENTRY and BIRTHS series are much more volatile than RGDP and their fluctuations are larger and shorter in time than those of the RGDP. Similar considerations hold for EXIT and DEATHS. Furthermore, notice that both the ENTRY and BIRTHS series seems to lag the RGDP in the recession of 2001, they seems instead coincident in the last recession of 2008. EXIT and DEATHS seem instead to lead the RGDP, at least in the last recession.

To further investigate the leading or lagging behavior of the disaggregated series we now compute the maximum absolute value cross-correlations between the cycle of the disaggregated ENTRY and EXIT and that of the RGDP. For completeness we consider the full sample of the disaggregated series, i.e. 1977–2013. When taking into account the nature of the indicators we are working with and their cyclical movements previously described, it seems reasonable to limit our search for the corresponding maximum absolute cross-correlation to a range between lags and leads of 6 quarters for the ENTRY and EXIT series[8]. According to the results reported in Table 4 the disaggregated ENTRY series from BDS is generally a lag and procyclical

[8]Notice that, whenever it was not possible to find a maximum inside this range we had to enlarge it. Our choice is motivated to avoid finding spurious results stemming from the fact that whenever the duration of the business cycles is short, a variable that leads (lags) the reference cycle by several months can be wrongly classified as lagging (leading) since it can be closer to the previous cycle than to the next. See also Altissimo *et al.* (2001, 2010) for a discussion of this problem. Looking at Figure 4 it is easy to see that the cycle of the ENTRY series is shorter than the cycle of Establishments EXIT. This motivates our choice of the different number of leads/lags for the two disaggregated series.

indicator of the business cycle with maximum absolute cross-correlation at the lag 3 for both the BK and the HP filter. On the contrary, both the univariate and the combined EXIT series are negatively correlated with RGDP and leads the cycle with a maximum cross-correlation at lag 6 for both the BK and the HP cycles. Also the contemporaneous correlations with RGDP are positive for ENTRY and negative for EXIT. However, they are not statistically significant.

Different story holds for the two BLS of BIRTHS and DEATHS, which shares the common feature of being generally pro-cyclical in the case of BIRTHS and counter-cyclical in the case of DEATHS. Namely, BIRTHS seems to be a coincident indicators, even though the correlation at lag 1 is almost equal to that at lag 0, while from lag 2 it decreases. The series of DEATHS is instead difficult to define since both the correlation at a lead and a lag 3 and 4 are high and significant, both for the BK and the HP filter respectively. Their negative correlation with RGDP at time zero and for all other leads, together with the economic intuition, however suggest to consider it as leading indicator and negatively correlated with RGDP. The SVAR analysis that follows will confirm this result, at least for the response of the EXIT and the DEATHS series to a TFP shock. Finally, it is important to notice that the high volatility of BDS-BLS series and its shorter cycle would probably make more appropriate a further analysis with equivalent monthly disaggregated series.

6 A Structural VAR Analysis

Once the statistical properties of the new disaggregated series has been investigated, it is possible to use them for structural analysis. To this aim, we run seven SVAR models for different combination of Entry and Exit proxies; see Table 7 for the description of the samples and system of each model, which are labeled as M1–M7. Then we compute the resulting Impulse Response Functions (IRFs) of all the system to a one standard deviation TFP shock[9]. All series in all systems are in logarithmic differences. For all the models considered we estimate a VAR(1) with a linear trend. Table 5 reports the estimates of M1, jointly with the results of standard diagnostic tests. The good properties of the fitted model can be appreciated in Figure 7, where a residual analysis is performed. In particular, for each variable, we plot the fitted value of OLS estimator (first column), their standardized residuals (second column), their estimated histogram and density jointly with the standard Normal case (third column), and the QQ plot (fourth column). The same considerations hold for the estimation of M2; see Figure 7, panel (b) and Table 5. Thus, we are confident in the feasibility of the M1 and M2 to serve as basis for IRF analysis. To identify the TFP shock we use a short-run restriction. This implies a standard identification based on Cholesky decomposition with the following ordering:

1. $DLTFP_t$,
2. $DLRGDP_t$,
3. $DLAVEntry_t$,
4. $DLAVExit_t$.

This allows to consider the technology shock as the most exogenous. The resulting IRFs, jointly with bootstrap confidence bands resulting from 1,000 draws, are plotted in Figure 11 – 17. Notice that, in both models the RGDP increases in response to the shock, firms ENTRY is pro-cyclical and persistent. EXIT is instead countercyclical in the medium run, while it reacts positively on impact, even though according to the confidence bands this positive impact response might be not statistically significant.

In general, responses of RGDP are almost immediate (between 1 and 3 quarters)

[9]Here, the TFP here used is the "Aggregate Productivity" series by Fernald and Matoba (2009). It can be downloaded, jointly with other related measures of productivity, from the websites of FED S.Francisco.

and of TFP just quite less (between 3 and 5 quarters). On the other hand, ENTRY and EXIT show a more persistent and hump-shaped pattern. According to the theory, one should expect a positive (negative) response for ENTRY (EXIT); this holds for M1 (where univariate UCM series are employed), as well as for M2 (Combined UCM), where indeed the ENTRY series is pro-cyclical and the EXIT series is countercyclical. M3 and M4 show the VAR system with the BLS and BED series - respectively of OPENINGS and CLOSINGS and BIRTHS and DEATHS. Their residual analysis is reported in figures 8, panel (a) and (b), respectively. While M3 passes substantially the diagnostic check, the same cannot be said for M4, where the histogram and density of the residual is dramatically distant from the assumed standard Normal; not surprisingly, the QQ plot of the estimated errors is almost constantly in the neighborhood of the bounds. Their IRFs are quite similar, albeit M3 is considerably more moderate in the magnitudo of the shock - approximately one half than M4.

What happen if considering our disaggregated data to the same time-span of BLS? M5 and M6 clearly shows that the responses of Entry and Exit are still consistent with the theory and now very similar to the responses of M1. The residual analysis of these VARs is reported in figures 9, panel (a) and (b), respectively. The IRFs are still more persistent. Moreover, EXIT presents an overshooting behavior, particularly evident in the Exit series of M6. The latter result on the overshooting of the exit series is consistent to what found theoretically by Rossi (2015).

Finally, the evidence of the residual and IRF analysis of M7 (in Figure 10 and 17, respectively) suggests that data from Economagic lead to a VAR system statistically poor and not consistent with the economic theory.

7 Conclusions

Recent advances in macroeconomic theory make the availability of new time series data on firms' dynamics a hot issue. This need is here satisfied by applying a new unobserved component-based temporal disaggregation method to the Census Bureau Research Data Center with emphasis on entry and exit of firms at establishment level.

Several conclusions descend from the results of our disaggregation and business cycle analysis. From a qualitative point of view, these series confirm the dynamics implied by theoretical models recently introduced in the literature. In facts, our SVAR models with the new quarterly BDS series on firm dynamics shows that a TFP shock is associated with a negative and persistent response of EXIT and a positive and persistent response of ENTRY. A further investigation of the responses of these variables to other demand and supply shocks is also needed and it is already in our research agenda.

From a methodological point of view, the time series derived by UCM are preferable to the CL-NM, being the ENTRY and EXIT unable to satisfy the theoretical condition of perfect cointegration between indicators (or, in case of single indicator, with original process). Hence, the need to repeat the UCM disaggregation for more, different indicators. However, the similarity of the combination of 134 disaggregated series deriving from the FRED-MD dataset with the ones resulting from univariate UCM poses a doubt on the effectiveness of this last *ad hoc* strategy. Such a result can be explained from the simple fact that many of the variables in the dataset have paths mutually different. This somehow compensates many of the potential differences in series singularly derived from these indicators. Thus, we recommend caution in the use of on the disaggregation methods based on large datasets without having a proper, economically meaningful selection of the indicators. Moreover, we are aware that using more properly multivariate methods can improve the disaggregation results. A deeper empirical and methodological investigation in this sense is highly recommended.

References

Altissimo, F., Bassanetti, A., Cristaldoro, R., Forni, M., Lippi, M., Reichlin, L. and Veronese, G. (2001). EUROCOIN: A Real Time Coincident Indicator of the Euro Area Business Cycle. CEPR Working Paper No. 3108.

—, Cristaldoro, R., Forni, M., Lippi, M., Reichlin, L. and Veronese, G. (2010). New EUROCOIN: Tracking Economic Growth in Real Time. *The Review of Economics and Statistics*, **92**, 1024–1034.

Bergin, P. and Corsetti, G. (2008). The extensive margin and monetary policy. *Journal of Monetary Economics*, **55**, 1222–1237.

—, Feng, L. and Lin, C. (2014). Financial Frictions and Firms Dynamics. NBER Working Paper no 20099.

Bilbiie, F., Ghironi, F. and Melitz, M. (2012). Endogenous Entry, Product Variety, and Business Cycle. *Journal of Political Economy*, **120**, 304–345.

Casares, M. and Poutinau, J. (2014). A DSGE Model with Endogenous Entry and Exit. Carleton Economics Paper no. 14-06.

Chow, G. and Lin, A. (1971). Best Linear Unbiased Interpolation, distribution and extrapolation of time series by related series. *The review of Economics and Statistics*, **53**, 372–75.

Colciago, A. and Rossi, L. (2012). Firms Entry, Oligopolistic Competition and Labor Market Dynamics. DNB Working Paper (Bank of Netherland).

— and — (2015). Firms Entry, Oligopolistic Competition and Labor Market Dynamics. DNB Working Paper No. 465.

de Jong, P. (1989). Smoothing and interpolation with the state-space model. *Journal of American Statistical Association*, **84**, 1085–1088.

— (1991). The diffuse Kalman Filter. *Annals of Statistics*, **19**, 1073–1083.

Etro, F. and Colciago, A. (2010). Endogenous Martket Structure and the Business Cycle. *The Economic Journal*, **120**, 1201–1233.

Fernald, J. and Matoba, K. (2009). Growth Accounting, Potential Output and Current Recession. FRBSF Economic Letter No. 2009/26.

Hamano, M. and Zanetti, F. (2015). Endogenous Product Turnover and Macroeconomic Dynamics. Working Paper No. 759, University of Oxford.

Harvey, A. (1989). *Forecasting, Structural Time Series Model and the Kalman-Filter*. Cambridge, UK: Cambridge University Press.

Jaimovich, N. and Floetotto, M. (2008). Firm Dynamics, Mark-up Variations and the Business Cycle. *Journal of Monetary Economics*, **55**, 1238–1252.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econmetrica*, **59**, 1551–1580.

La Croce, C. and Rossi, L. (2015). Firms Endogenous Entry and Monopolistic Banking in a DSGE Model. *Macroeconomic Dynamics*, **Forthcoming**.

Lewis, V. (2009). Business Cycle Evidence on Firm Entry. *Macroeconomic Dynamics*, **127**, 324–348.

— (2013). Optimal Monetary Policy and Firm Entry. *Macroeconomic Dynamics*, **17**, 1687–1710.

— and Poilly, C. (2012). Firm entry, markups and the monetary transmission mechanism. *Journal of Monetary Economics*, **59**, 670–685.

McCraken, M. and Ng, S. (2015). Fred-md: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, **Accepted**.

Proietti, T. (2006). Temporal Disaggregation by state space methods: Dynamic regression methods revisited. *Econometric Journal*, **9**, 357–372.

Rossi, L. (2015). Endogenous Firms' Exit, Inefficient Banks and Business Cycle Dynamics. DEM Working Paper, University of Pavia.

Siemer, M. (2014). Firm Entry and Employment Dynamics in the Great Recession. FEDS Working Paper No. 2014-56.

A The Augmented Kalman Filter and Smoother

Using the initial conditions in (9), and further defining $A_1^* = W_1^*$, $P_1^* = H_1^* H_1^{*'}$, $q_1 = 0$, $s_1 = 0$, $S_1 = 0$, the augmented Kalman filter consists of the following equations and recursions: for $t = \tau s$, $\tau = 1, \ldots, [n/s]$ (that is, y_t^c available),

$$
\begin{aligned}
v_t &= y_t^c - z^{*'} a_t^{*'}, & V_t' &= z^{*'} A_t^*, & (17)\\
f_t &= z^{*'} P_t^* z^*, & K_t &= T_t^* P_t^* z^* / f_t, & (18)\\
\alpha_{t+1}^* &= T_{t+1}^* a_t^* + K_t v_t, & A_{t+1}^* &= W_{t+1}^* A_t^* T_{t+1}^* + K_t V_t' & (19)\\
P_{t+1}^* &= T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'} - K_t K_t' f_t & s_{t+1} &= s_t V_t v_t / f_t & (20)\\
q_{t+1} &= q_t + v_t^2 / f_t & d_{t+1} &= d_t + \ln f_t & (21)\\
S_{t+1} &= S_t + V_t V_t' / f_t & t &= 1, \ldots, T & (22)\\
& & & & (23)
\end{aligned}
$$

and, alternatively, for $t \neq \tau s$, (that is, y_t^c is missing),

$$
\begin{aligned}
\alpha_{t+1}^* &= T_{t+1}^* a_t^* & A_{t+1}^* &= W_{t+1}^* A_t^* T_{t+1}^* & (24)\\
P_{t+1}^* &= T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'} & s_{t+1} &= s_t & (25)\\
q_{t+1} &= q_t, & S_{t+1} &= S_t & (26)
\end{aligned}
$$

V_t' denotes a row vector with k elements. The quantities q_t, S_t, s_t accumulate weighted sum of squares and cross-products that will serve the estimation of β via generalized regression.

Notice that the quantities f_t, K_t (the Kalman gain) and P_t do not depend on the observations and that the first two are not computed when y_t^c is missing. Missing values imply that updating operations, related to the new information available, are skipped.

The augmented KF computes all the quantities that are necessary for the evaluation of the likelihood function. The filtered, or real time, estimates of the state vector and their estimation error matrix are computed as follows:

$$\hat{\alpha}_t|t = a_t^* - A_t^* S_t^{-1} s_t + P_t^* z^* \hat{v}_t / f_t, \qquad P_{t|t}^* = P_t^* + A_t^* S_t^{-1} A_t^{*'} - P_t^* z^* z^{'} P_t^* / f_t \quad (27)$$

The augmented smoothing algorithm proposed by de Jong (1989) can be appropriately adapted to hand missing values. Let define $r_T = 0$, $R_T = 0$, $N_T = 0$ and $\hat{v}_t = v_t - V_t^{'} S_t^{-1} s_t$. Then for $t = N, \ldots, 1$ and $t = \tau s$ (that is, y_t^c is available),

$$r_{t-1} = z^* v_t / f_t + (T_{t+1} + K_t z^{*'}) r_t, \qquad R_{t-1} = z^* V_t^{'} / f_t + (T_{t+1} - K_t z^{*'}) R_t \quad (28)$$

$$N_{t-1} = z^* z^{*'} / f_t + (T_{t+1} - K_t z^{*'}) N_t (T_{t+1} - K_t z^{*'})', \quad (29)$$

while, for $t \neq \tau s$ (that is, y_t^c missing),

$$r_{t-1} = T_{t+1} r_t, \qquad R_{t-1} = T_{t+1} R_t, \qquad R_{t-1} = T_{t+1} N_t T_{t+1}^{\prime}. \quad (30)$$

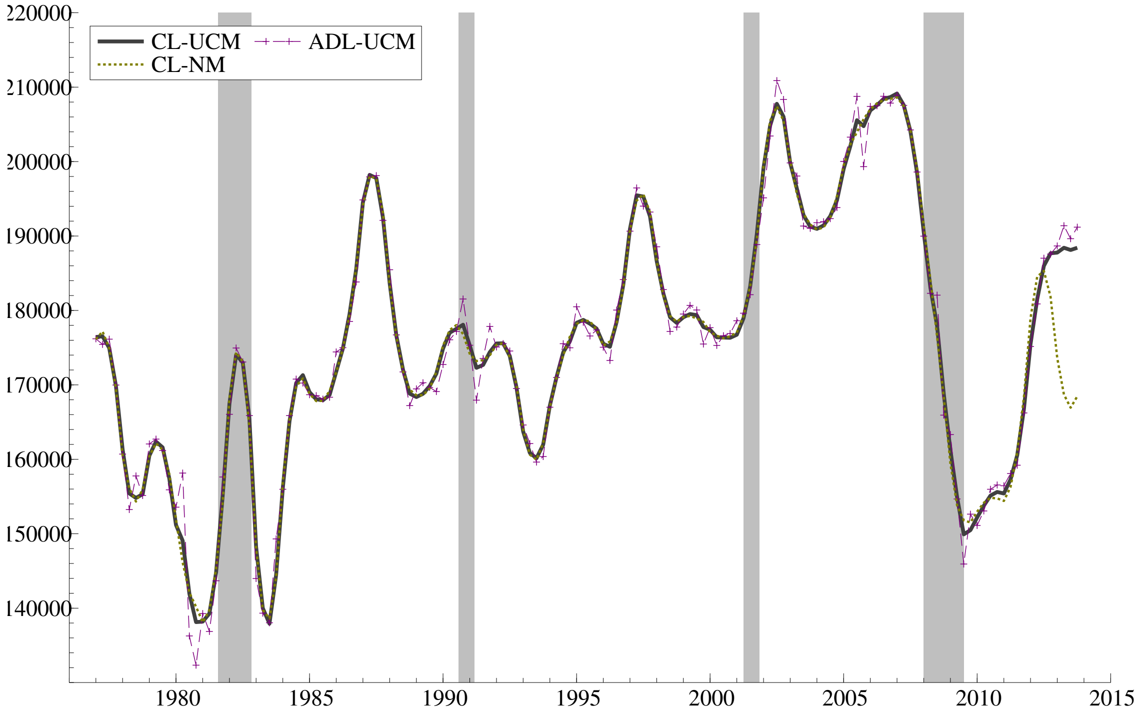
The smoothed estimates are obtained as

$$\hat{\alpha}_{t|t}^* = a_t^* + A_t^* \hat{\beta} + P_t^* (r_{t-1} + R_{t-1} \hat{\beta}) \qquad P_{t|t}^* = P_t^* + A_t^* S_{n+1}^{-1} A_t^{*'} - P_t^* N_{t-1} P_t^* \quad (31)$$

B Tables and Figures

Figure 1: Disaggregated series using different techniques

(a) Enstablishment Entry



(b) Enstablishment Exit

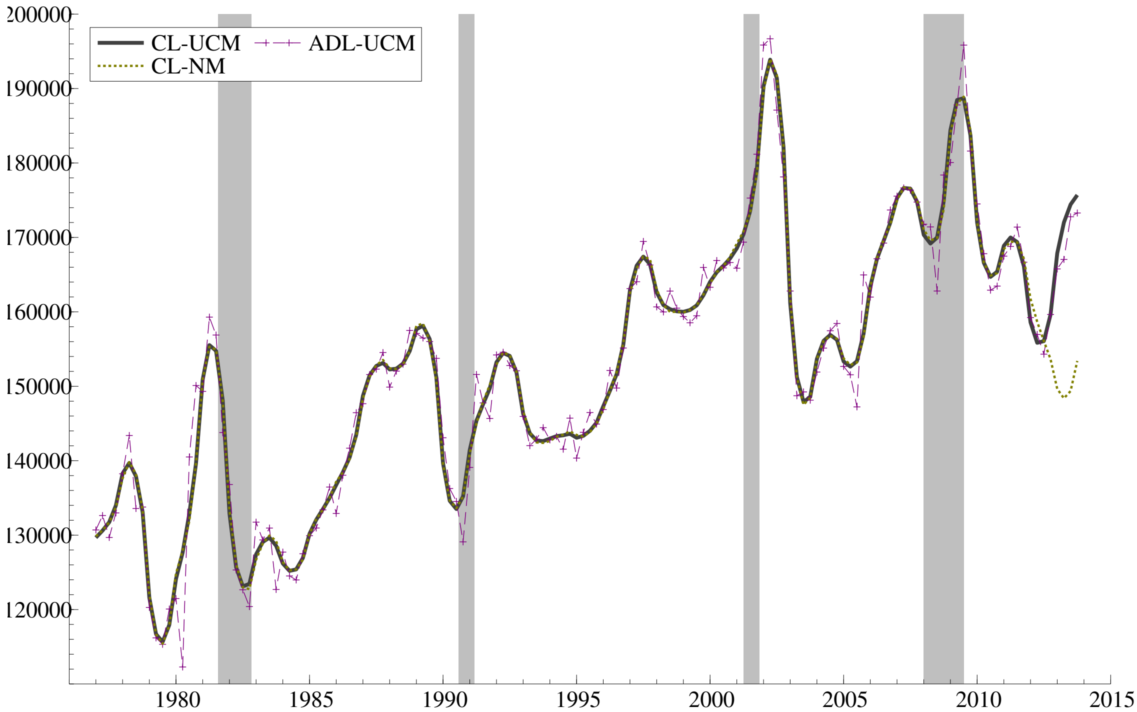
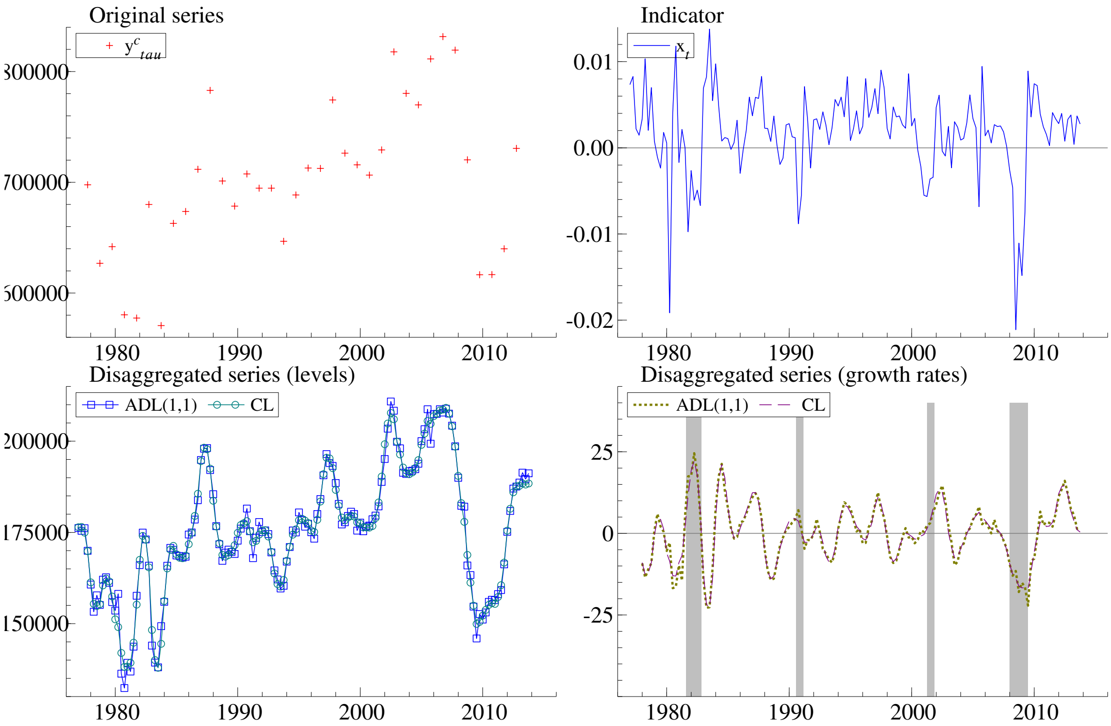


Figure 2: Comparison of different regression models in univariate Proietti Method

(a) Entry



(b) Exit

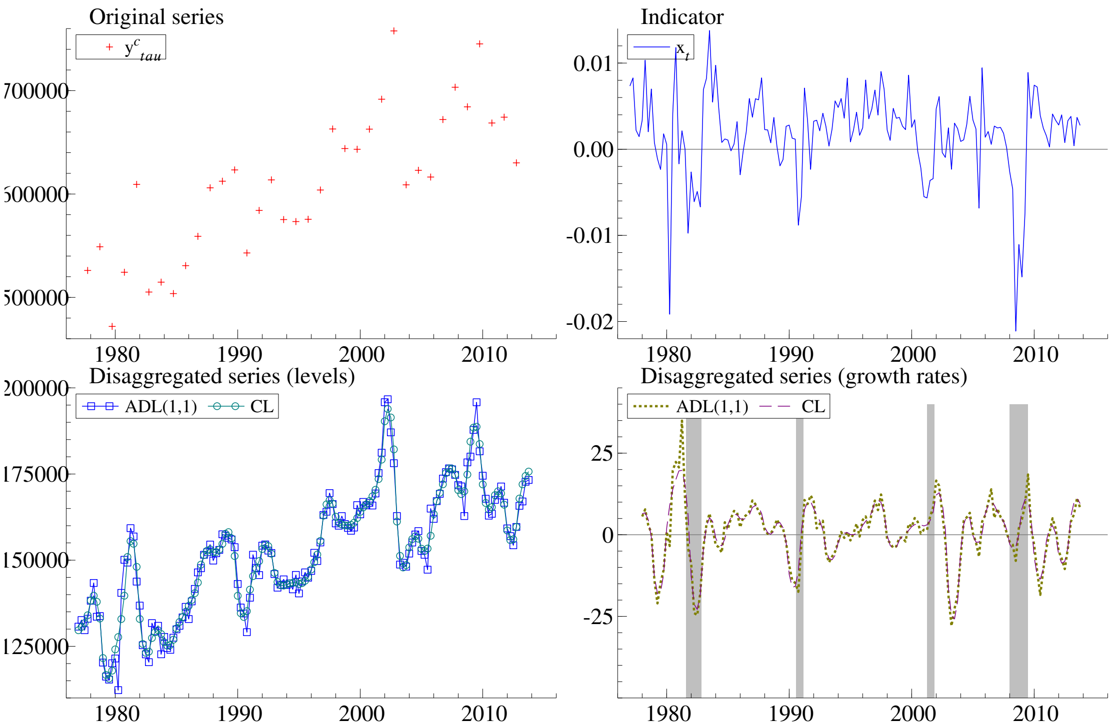
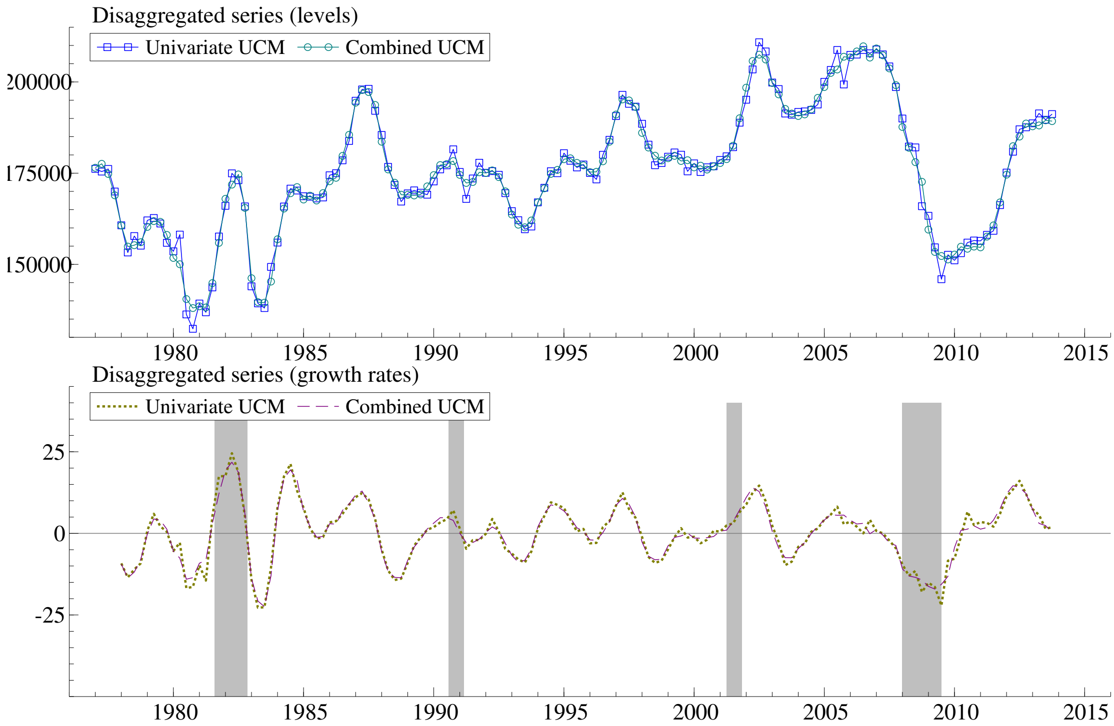


Figure 3: Comparison of univariate and combined UCM. Method

(a) Entry



(b) Exit

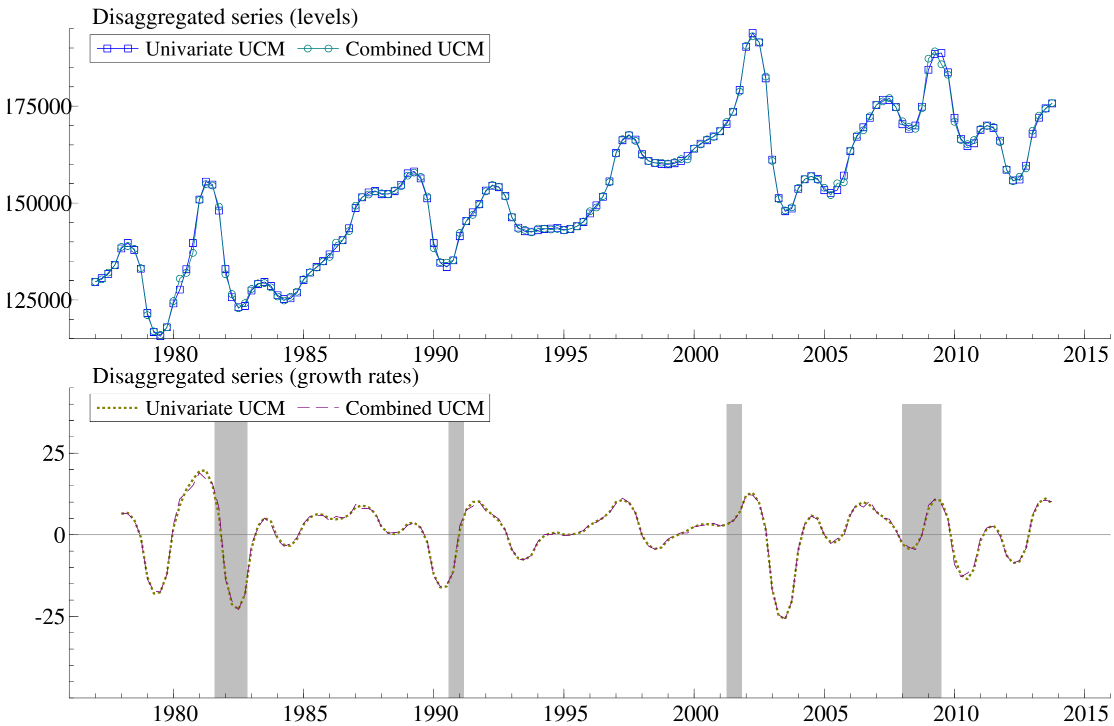
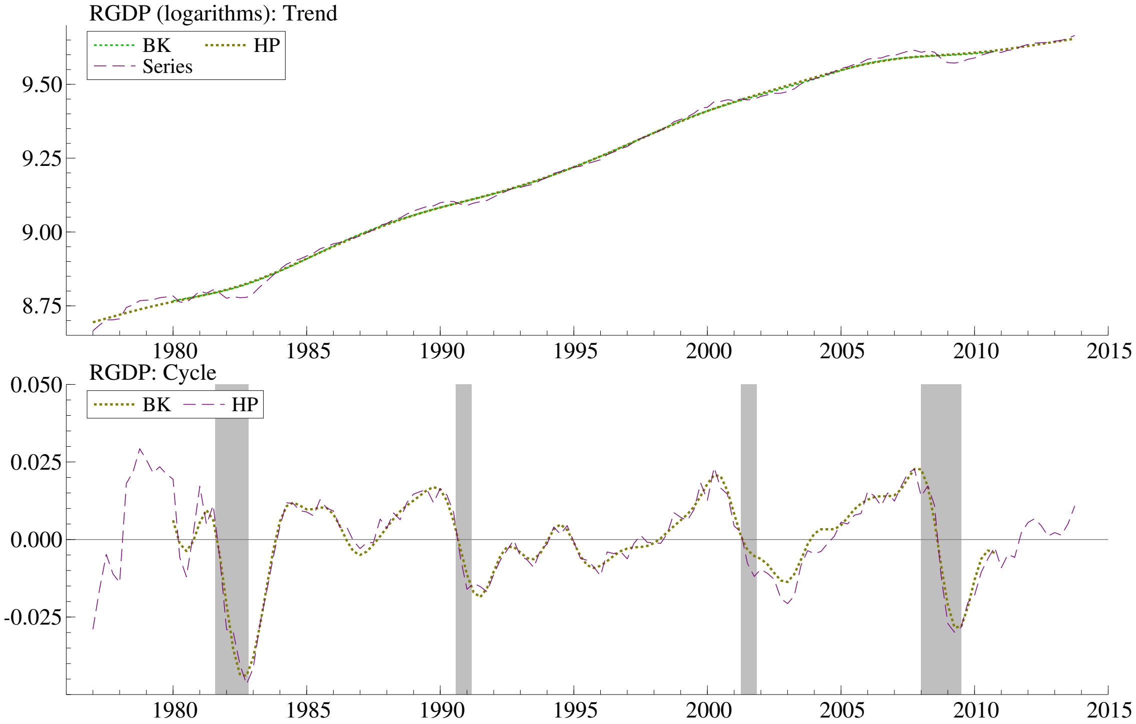


Figure 4: Business Cycle Analysis: RGDP and disaggregated BDS series

(a) RGDP



(b) Disaggregated BDS series

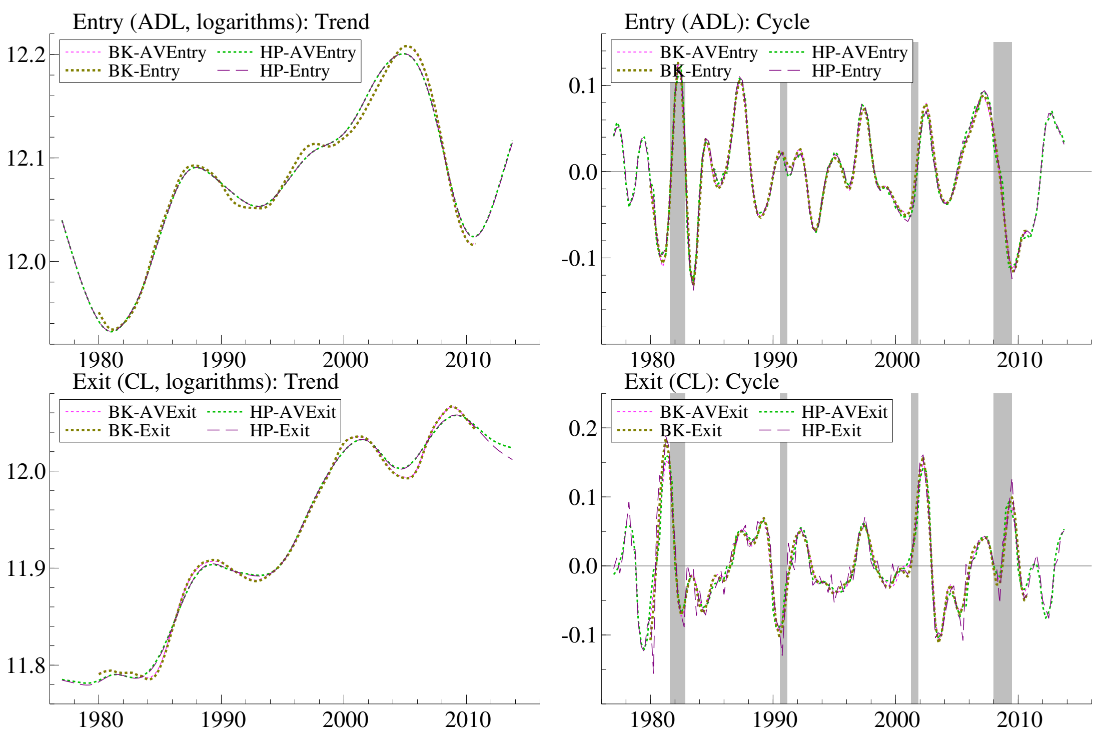
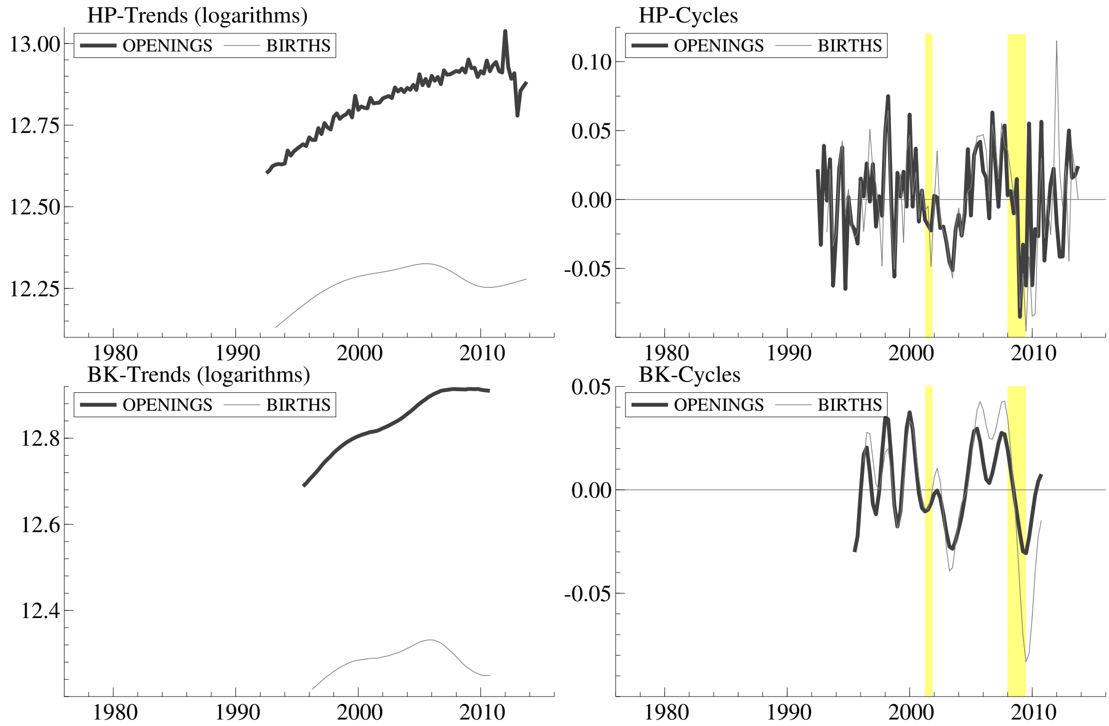


Figure 5: Business Cycle Analysis of BLS series

(a) Entry



(b) Exit

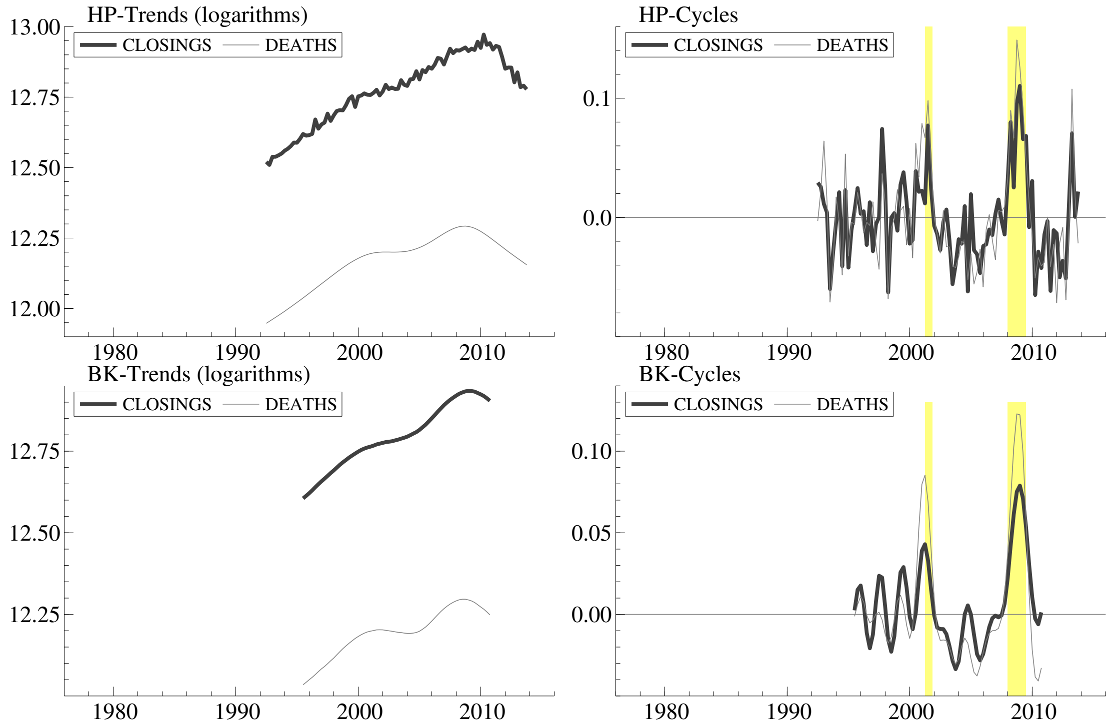
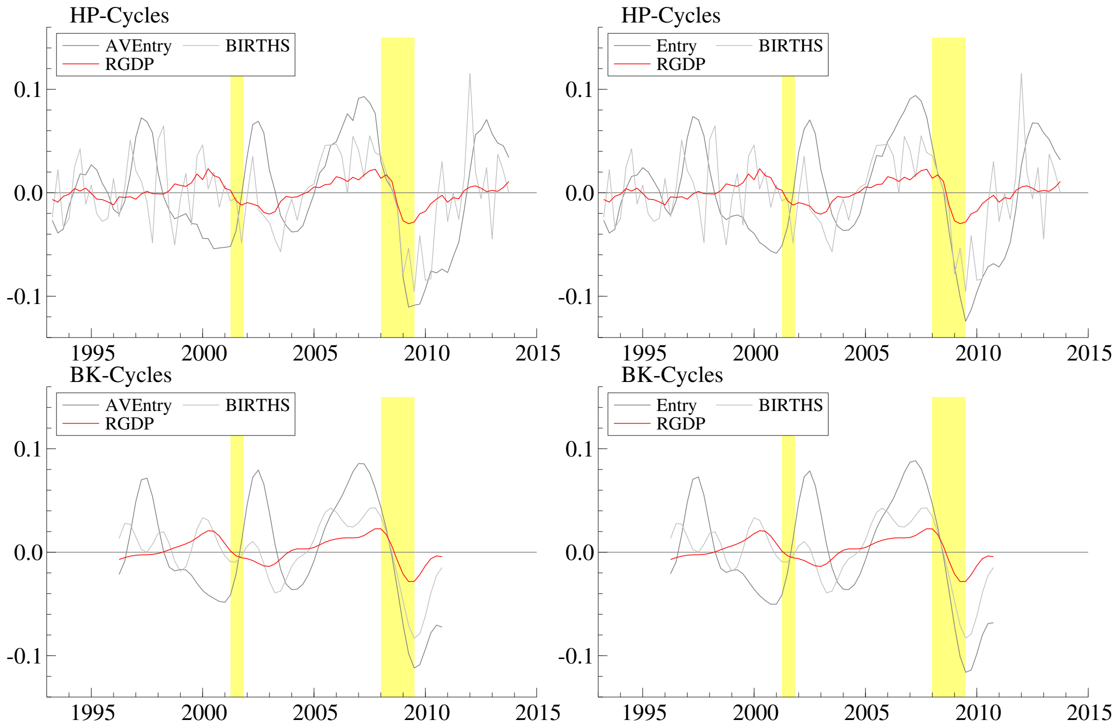


Figure 6: Business Cycle Analysis: comparison between proxies

(a) ENTRY



(b) EXIT

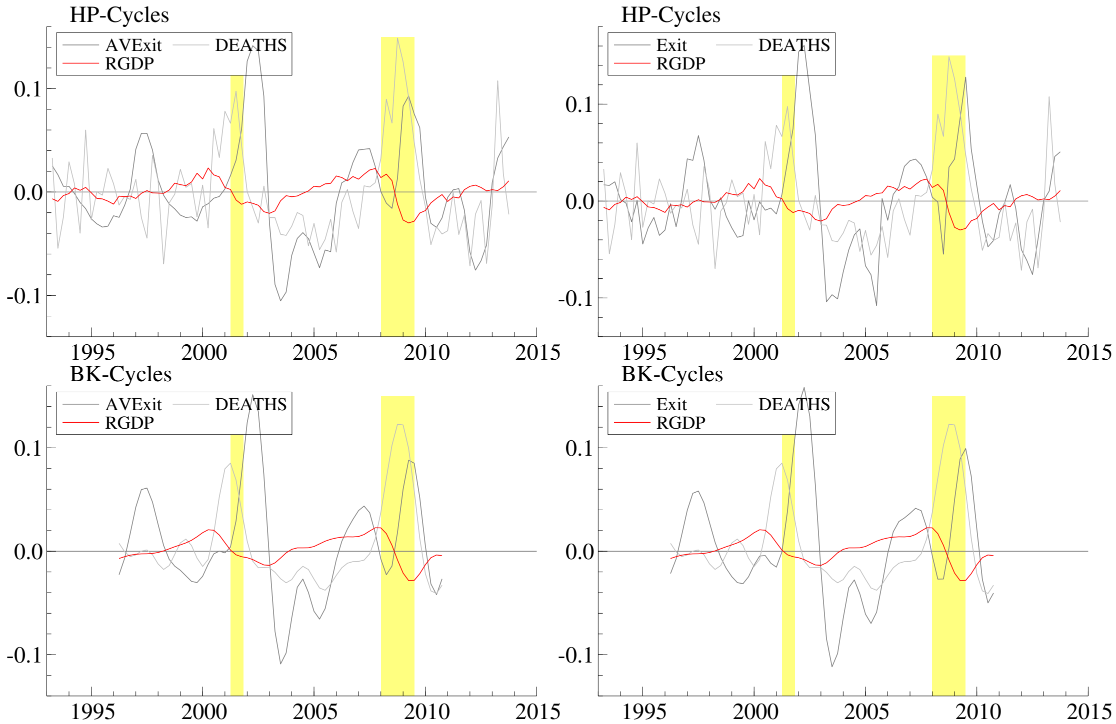
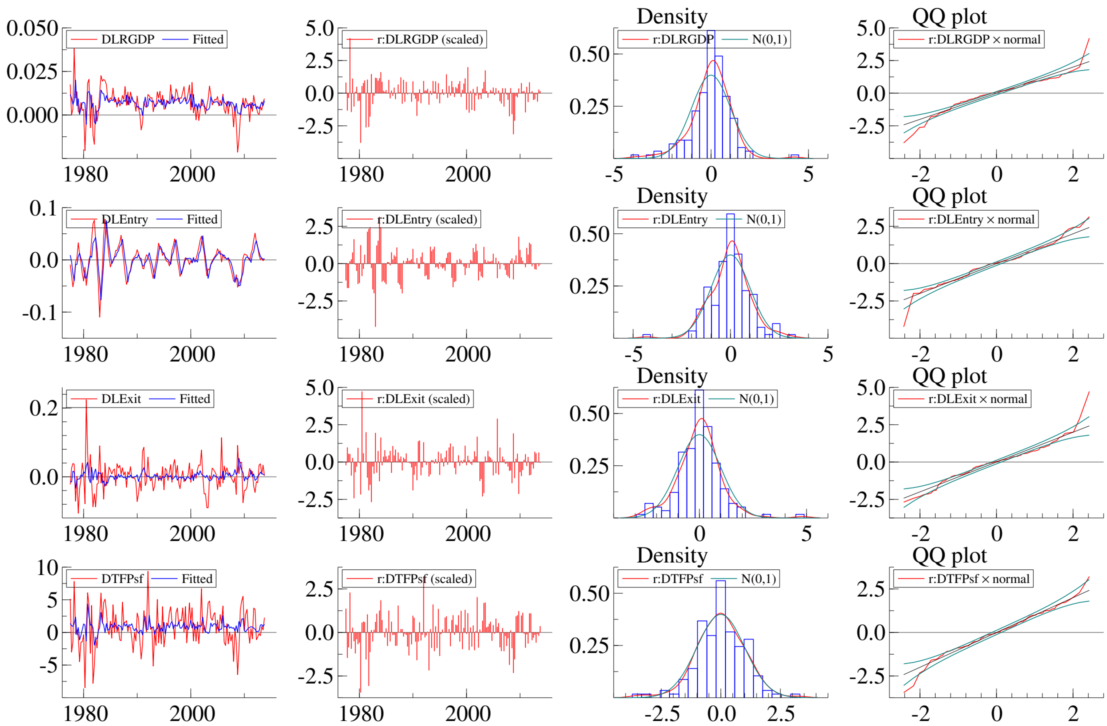


Figure 7: The estimated VAR models

(a) M1 (Univariate UCM, Sample: 1977:Q1–2013:Q4)



(b) M2 (Combined UCM, Sample: 1992:Q3–2013:Q4)

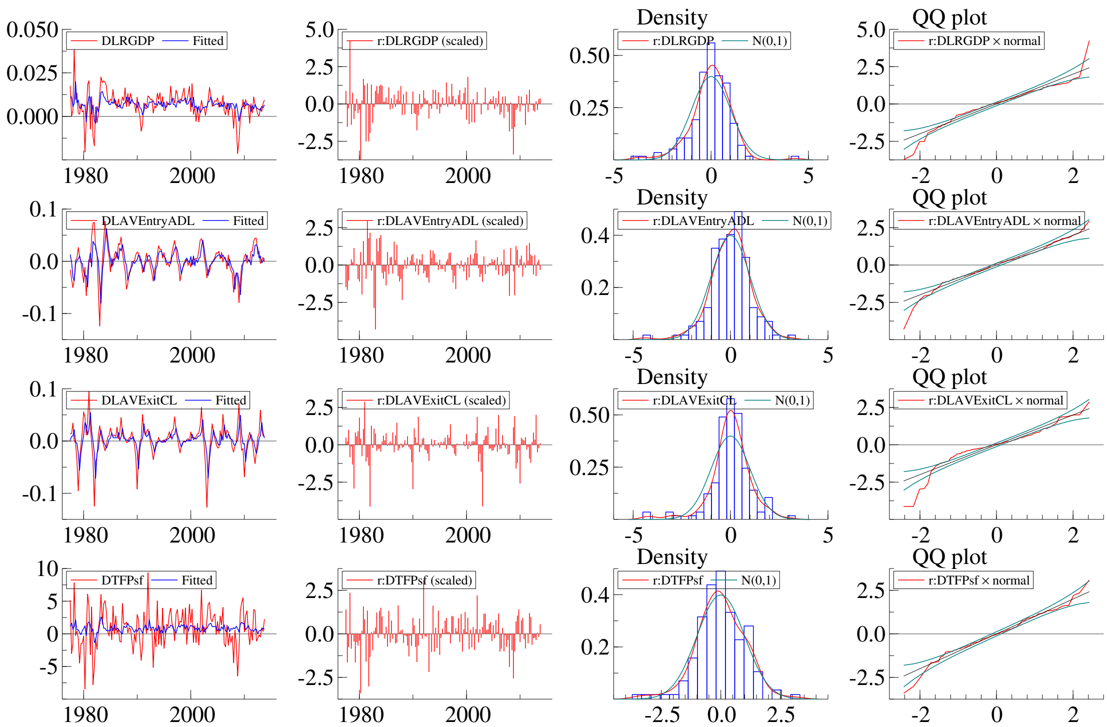
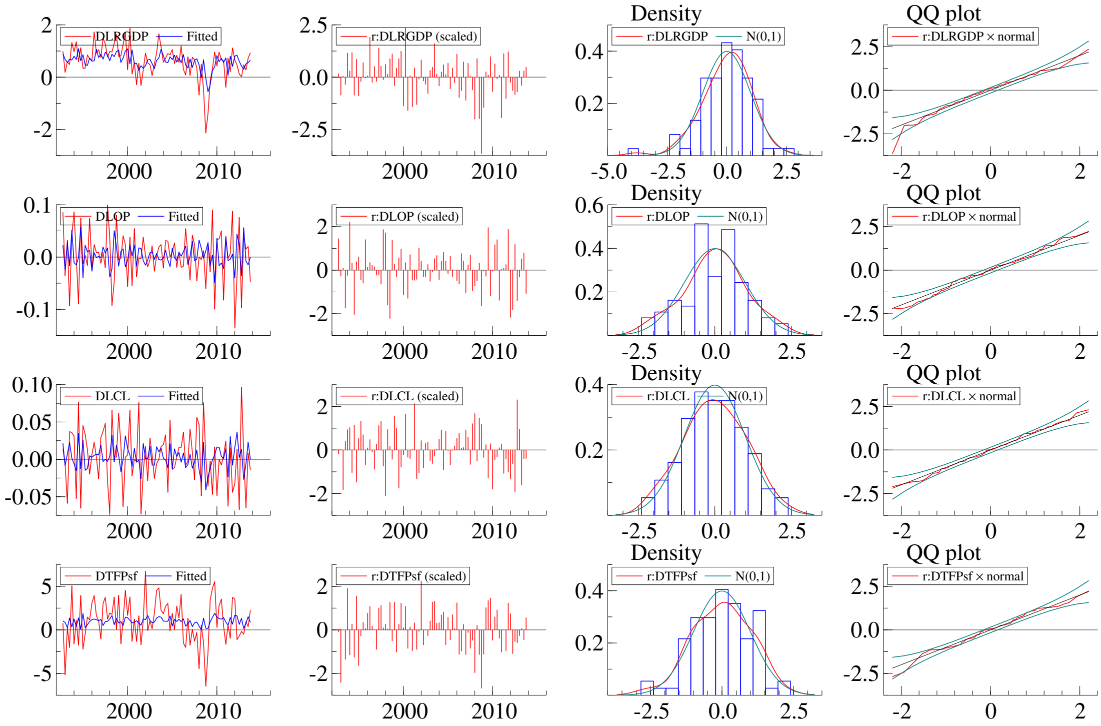


Figure 8: (...Continue)

(a) M3 (Data from BLS, Sample: 1992:Q3–2013:Q4)



(b) M4 (Data from QCEW, Sample:1993:Q2–2013:Q4)

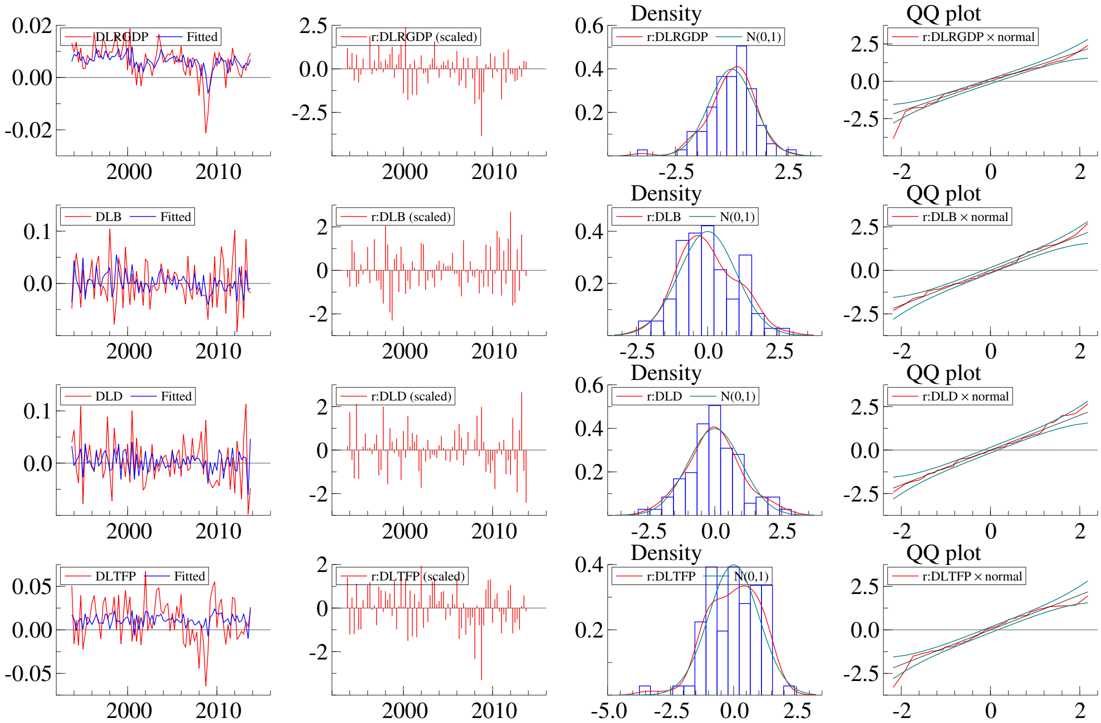
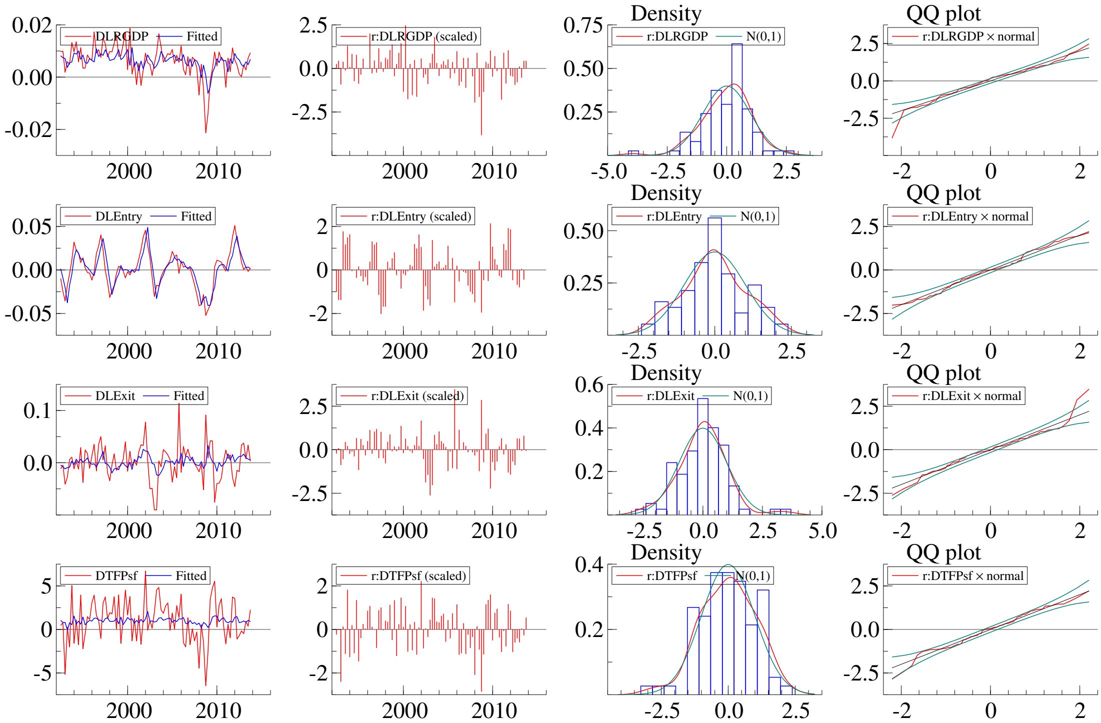


Figure 9: (...Continue)

(a) M5 (Univariate UCM, Sample: 1992:Q3–2013:Q4)



(b) M6 (Sample: 1992:Q3–2013:Q4)

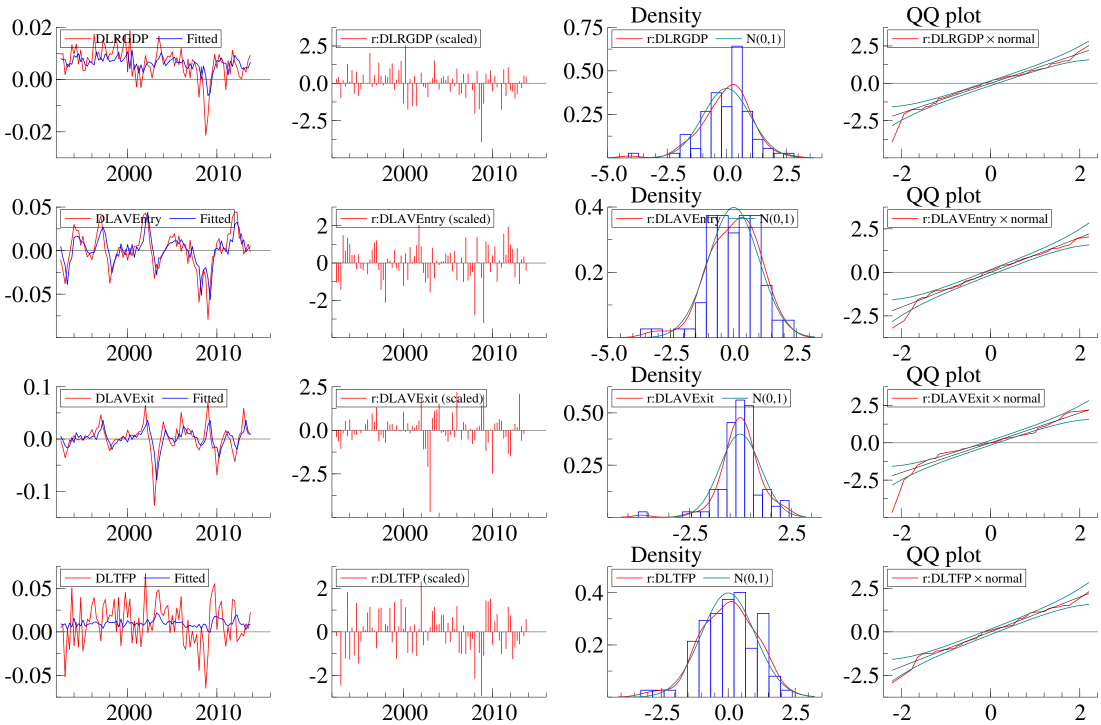


Figure 10: (...Continue)

(a) M7 (Data from Economagic, Sample: 1977:Q1–1996:Q3)

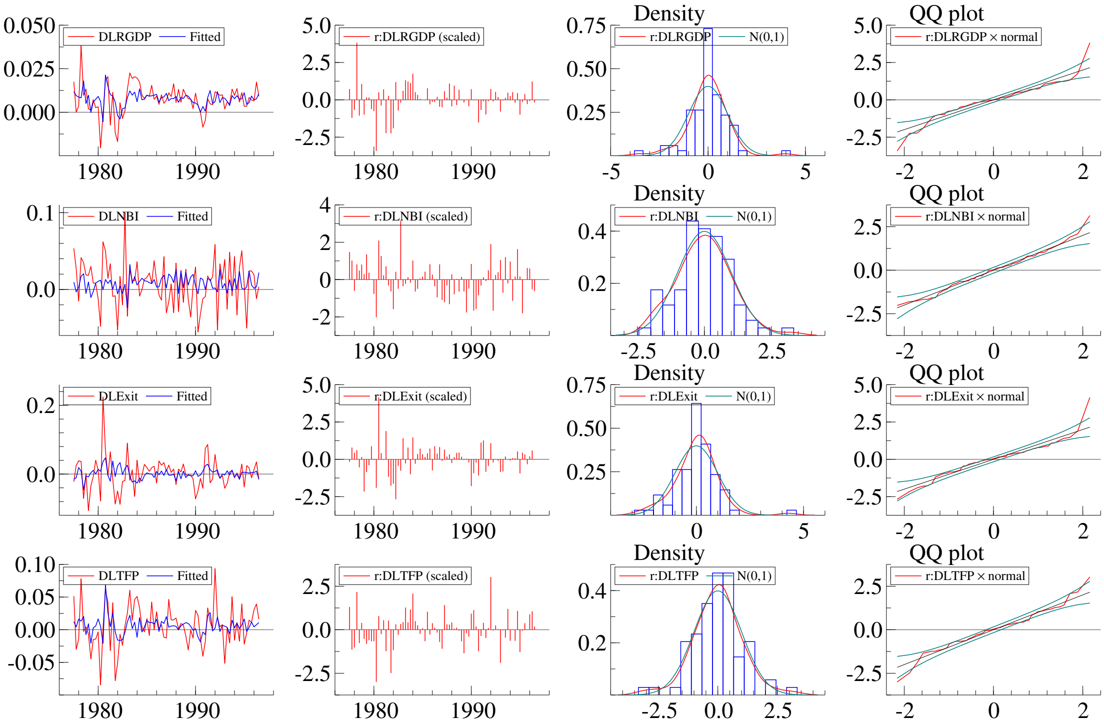


Figure 11: The Structural Analysis: M1 (Sample: 1977:Q1–2013:Q4, univariate UCM)

(a) TFP > TFP **(b)** RGDP > TFP

(c) Entry > TFP **(d)** Exit > TFP

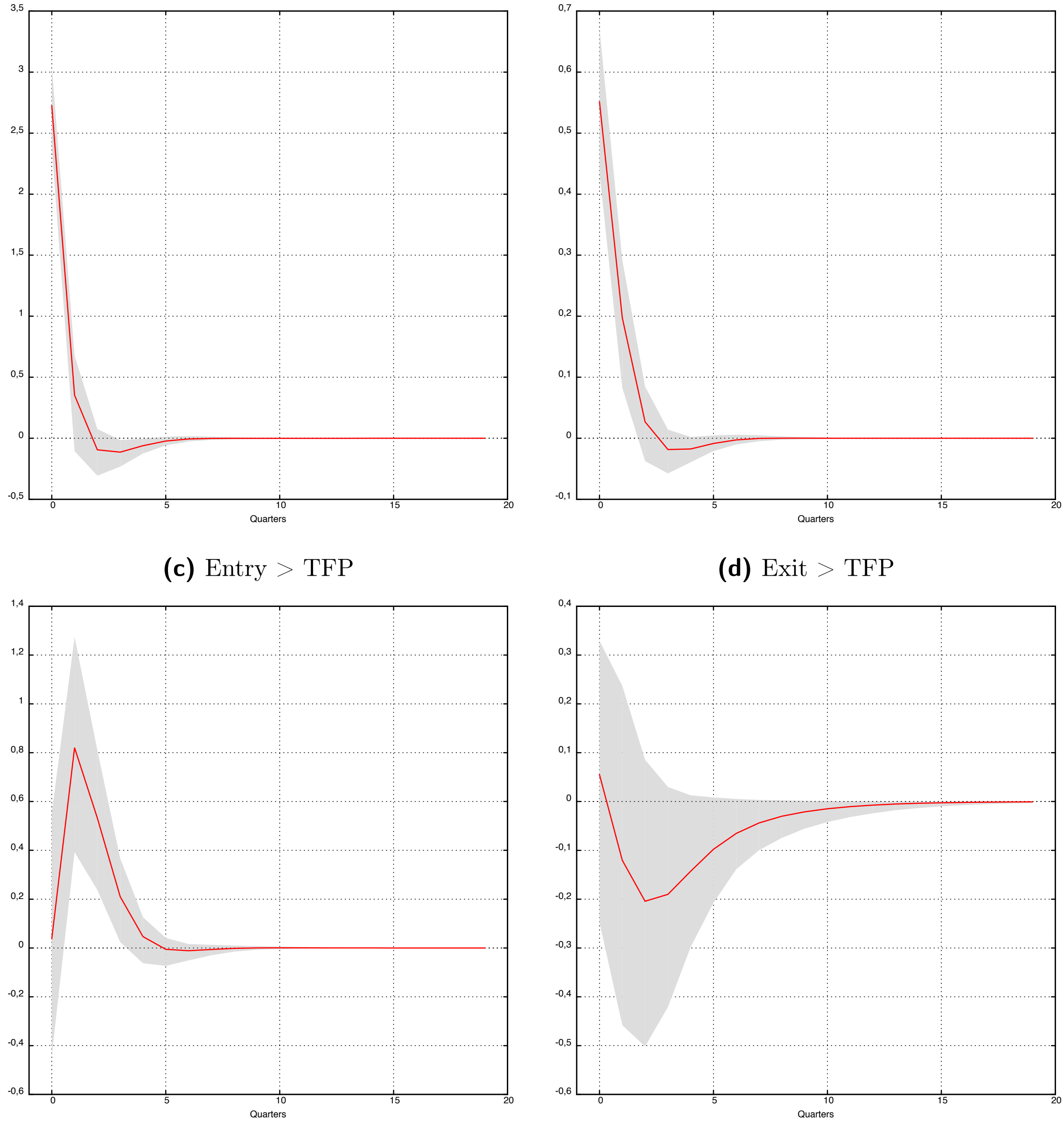
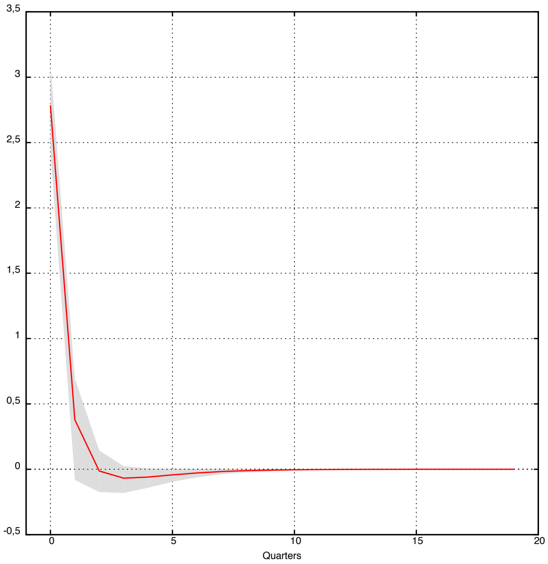
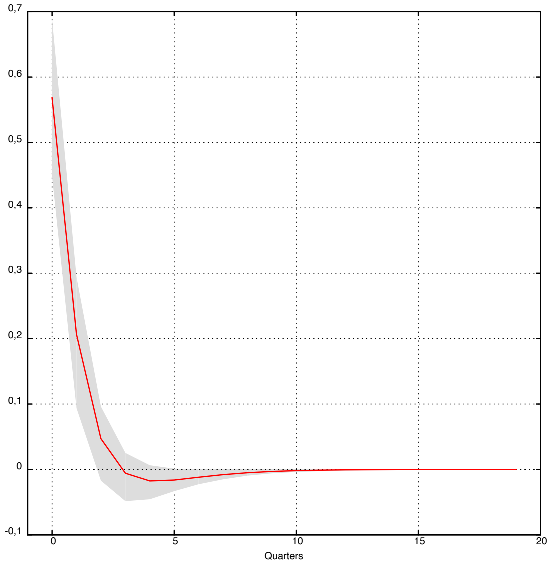


Figure 12: The Structural Analysis: M2 (Sample: 1977:Q1–2013:Q4, Combination UCM)

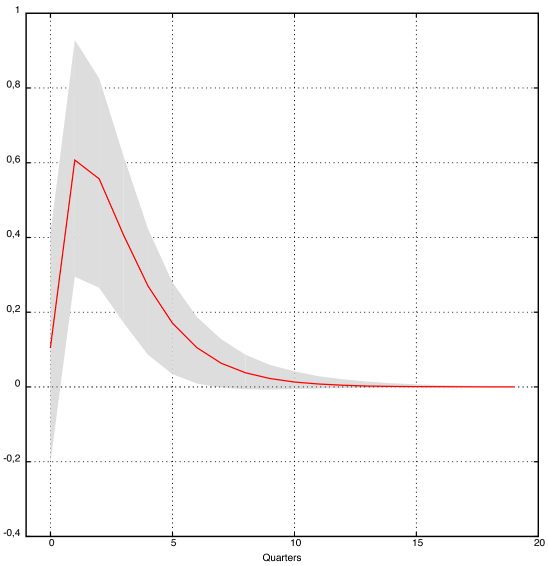
(a) TFP > TFP



(b) RGDP > TFP



(c) AVEntry > TFP



(d) AVExit > TFP

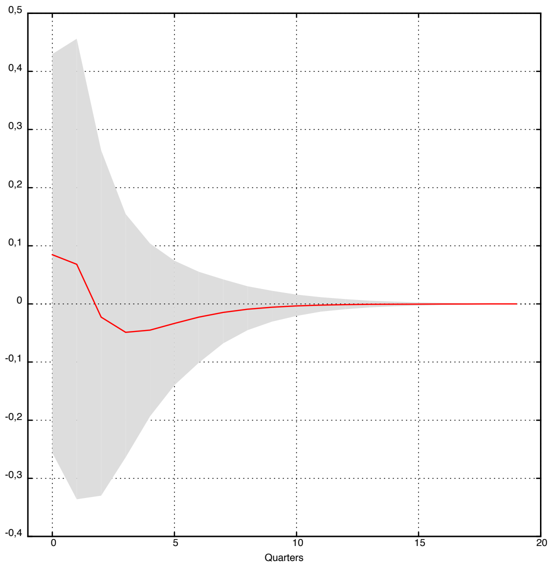
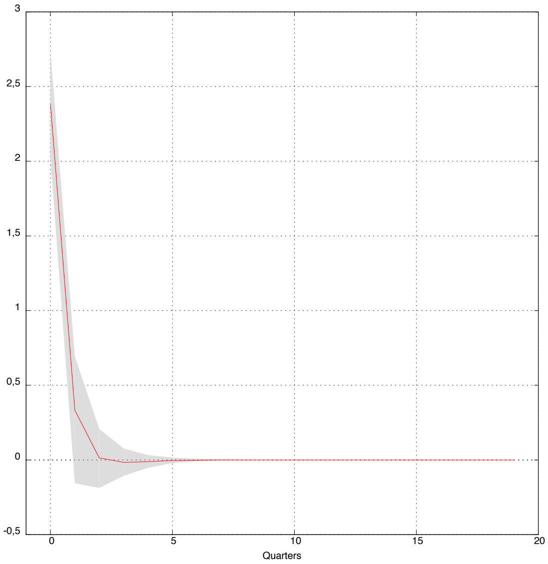
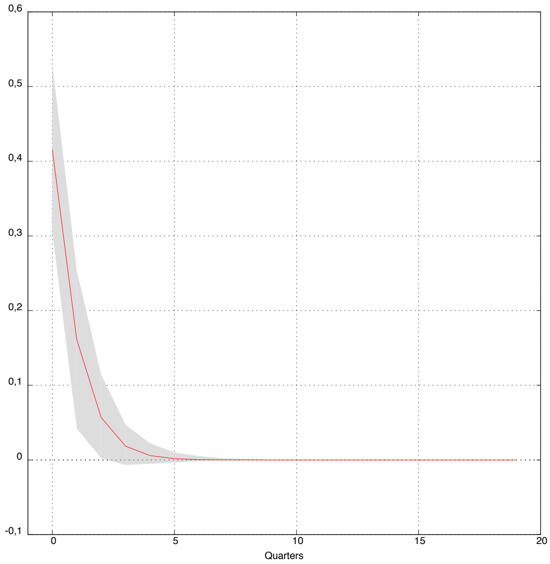


Figure 13: The Structural Analysis: M3 (Sample: 1992:Q3–2013:Q4, BLS Data)

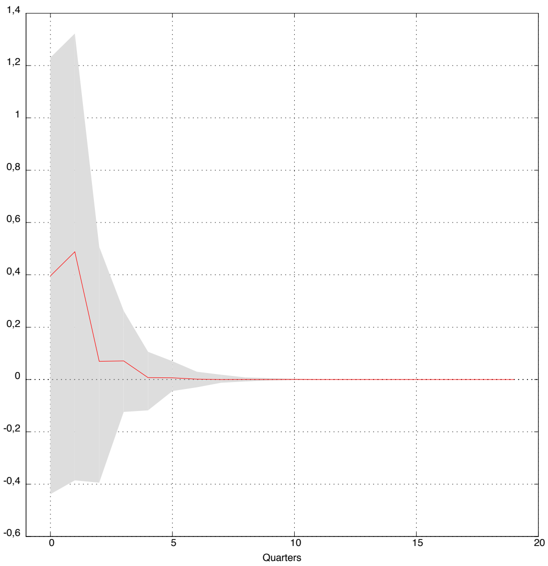
(a) TFP > TFP



(b) RGDP > TFP



(c) OPENINGS > TFP



(d) CLOSINGS > TFP

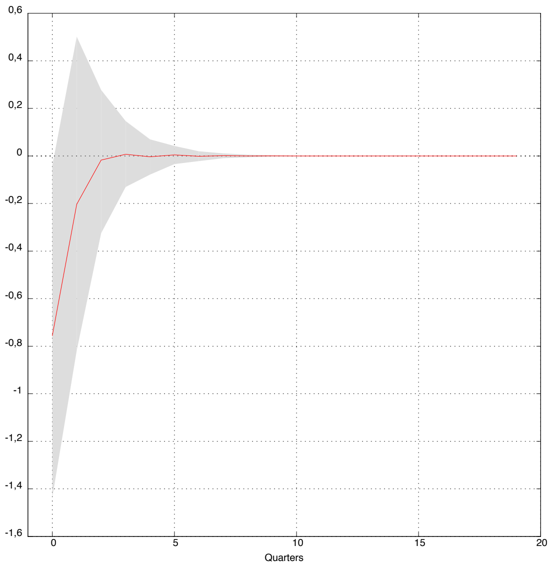
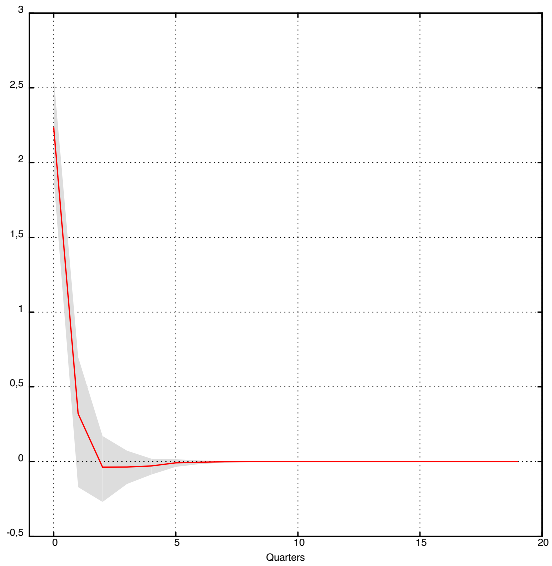
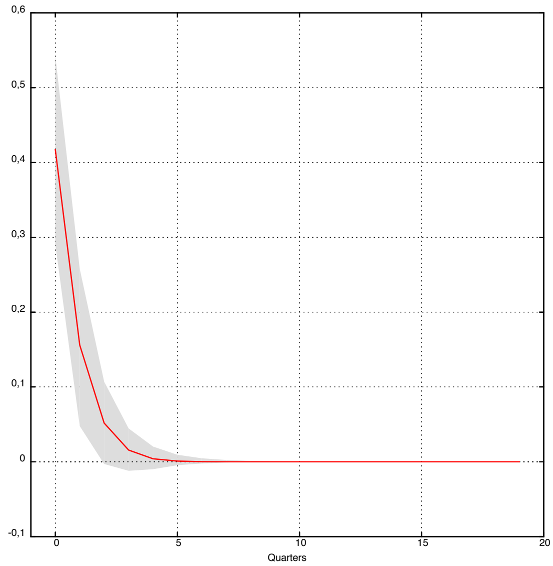


Figure 14: The Structural Analysis: M4 (Data by QCEW, Sample: 1993:Q2–2013:Q4)

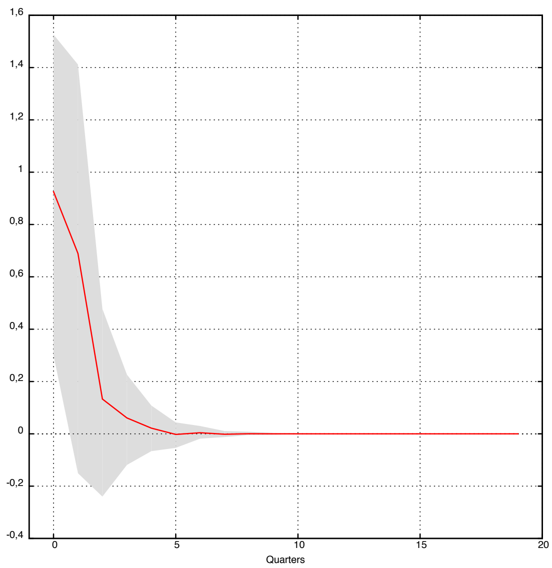
(a) TFP > TFP



(b) RGDP > TFP



(c) Births > TFP



(d) Deaths > TFP

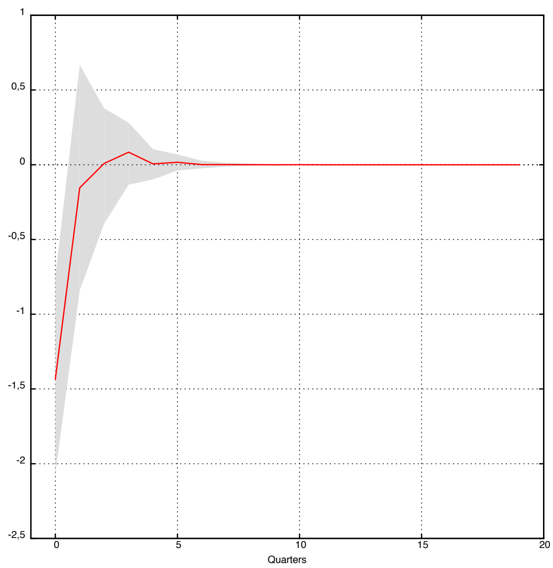
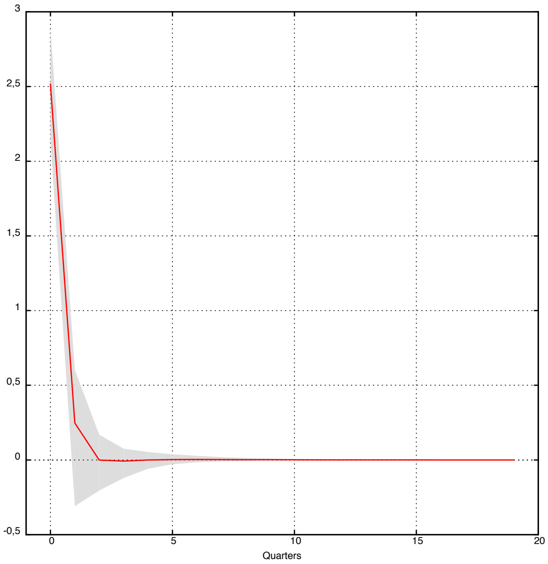
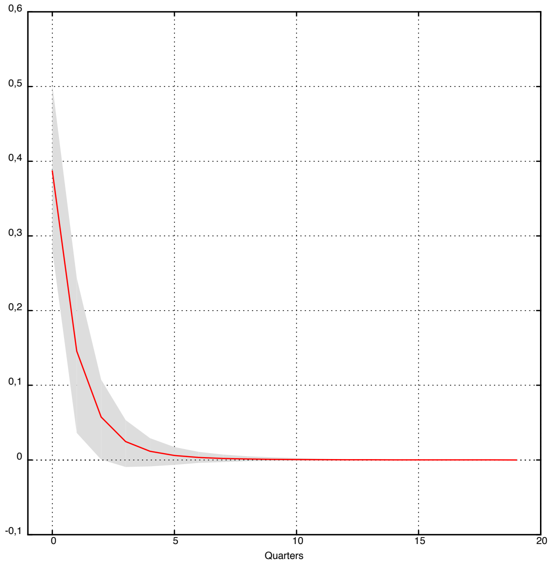


Figure 15: The Structural Analysis: M5 (Univariate UCM, Sample: 1992:Q3–2013:Q4)

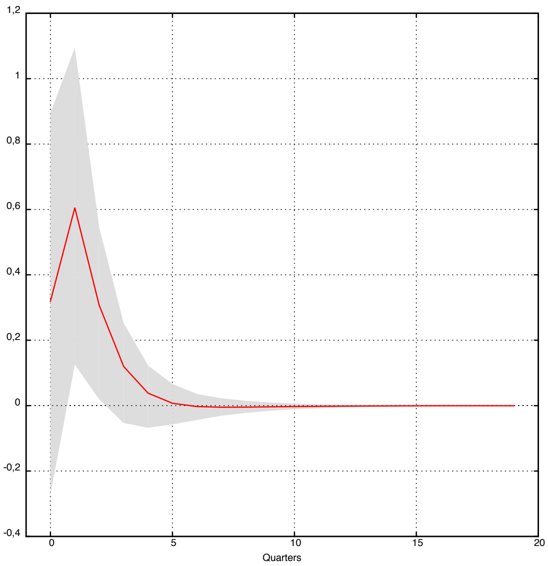
(a) TFP > TFP



(b) RGDP > TFP



(c) ENTRY > TFP



(d) EXIT > TFP

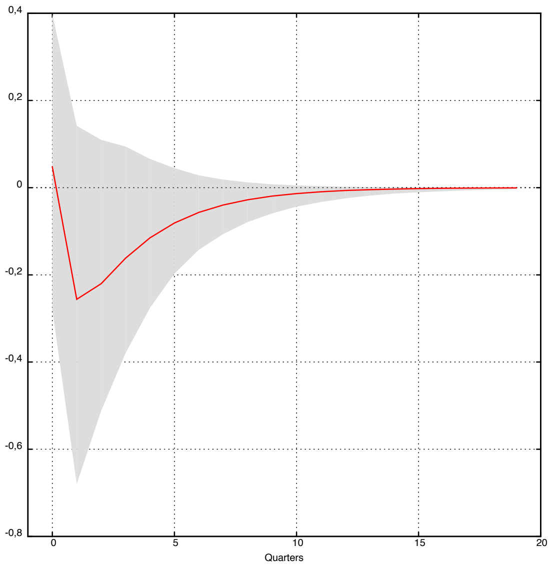
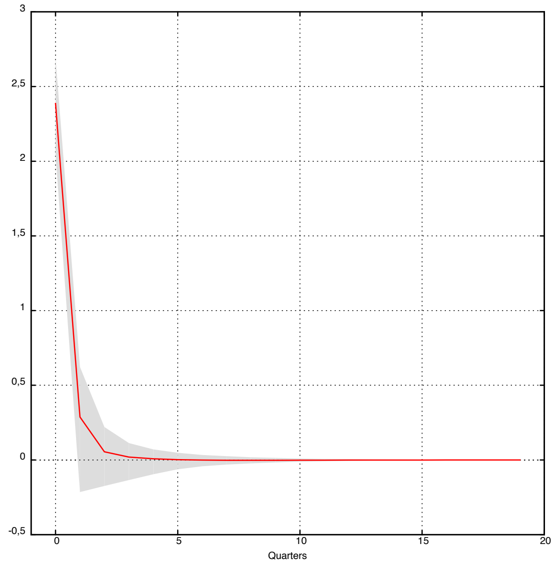
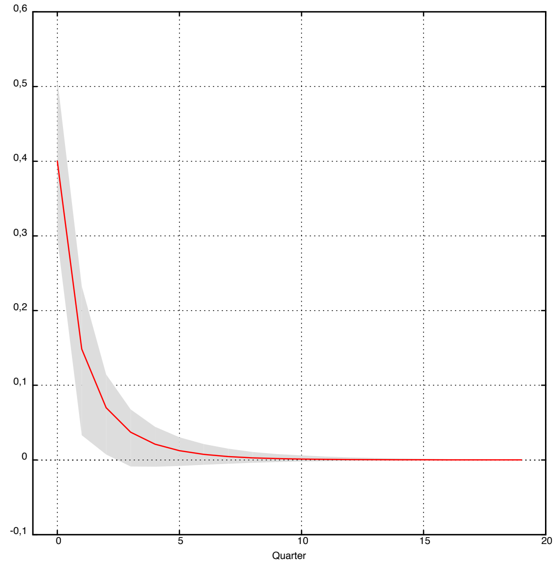


Figure 16: The Structural Analysis: M6 (Combination UCM, Sample: 1992:Q3–1996:Q3)

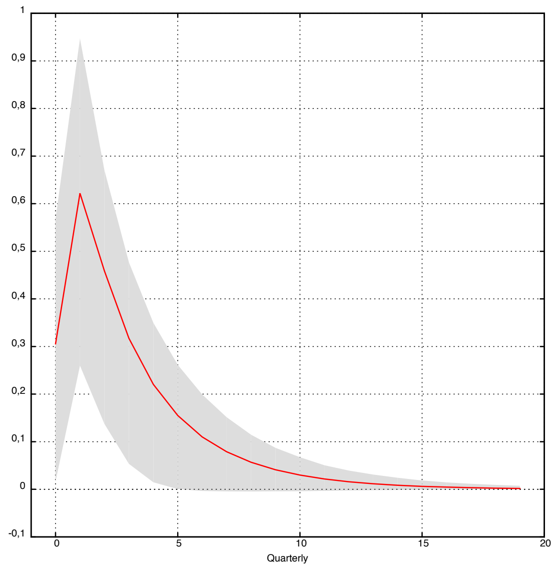
(a) TFP > TFP



(b) RGDP > TFP



(c) NBI > TFP



(d) EXIT > TFP

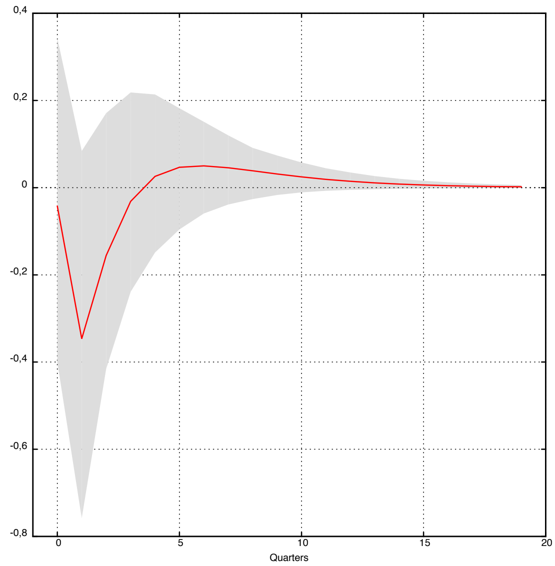
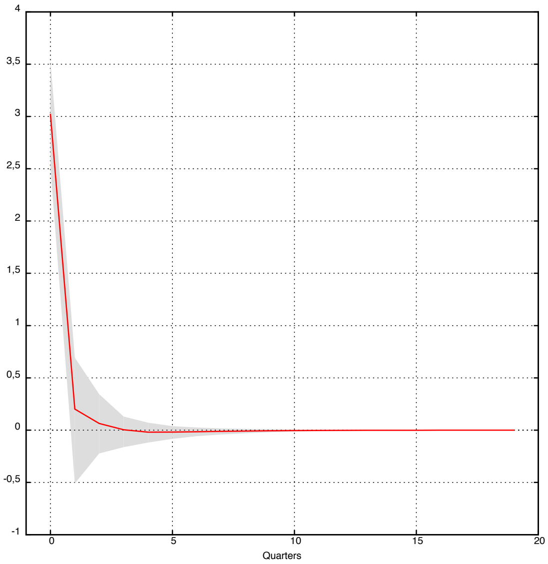
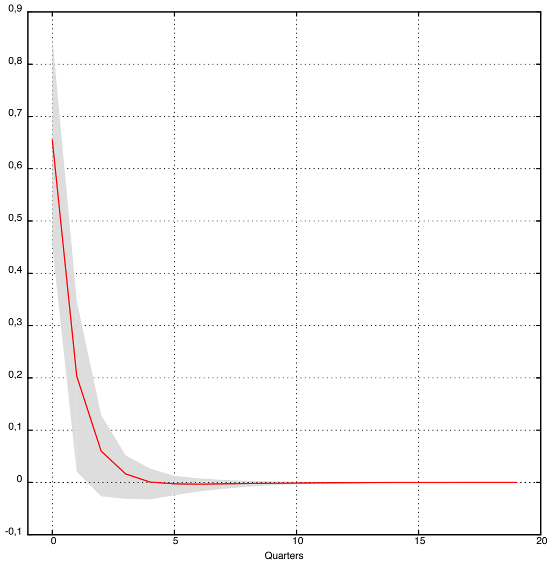


Figure 17: The Structural Analysis: M7 (Data from Economagic and univariate UCM, Sample: 1992:Q3–1996:Q3)

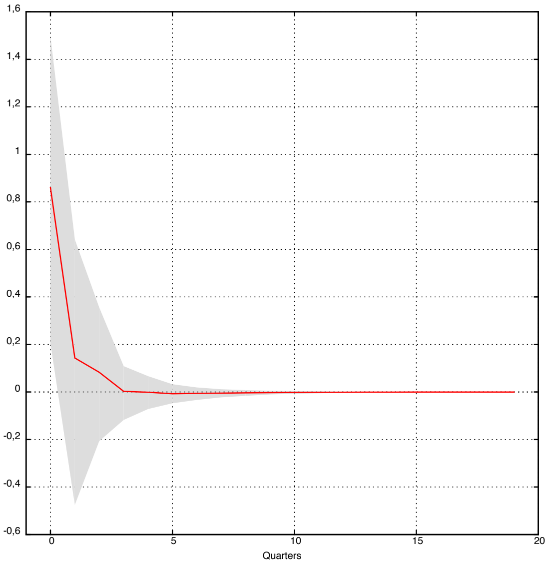
(a) TFP > TFP



(b) RGDP > TFP



(c) NBI > TFP



(d) EXIT > TFP

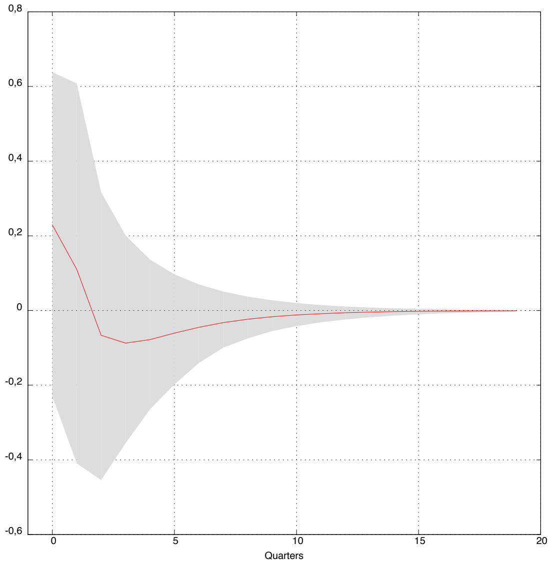


Table 1: Johansen Test

Rank	Trace statistic	$p-$value
0	177.101	0.000**
1	43.766	0.000**

Table 2: Correlations between disaggregated series with BLS data

Method	Corr(ENTRY, OPENINGS)	Corr(EXIT, CLOSINGS)	SAMPLE
CL-NM	0.1625	0.5279	
CL-UCM	0.2669	0.7171	1992:Q3–2013:Q4
ADL-UCM	0.2703	0.7225	
	Corr(ENTRY, BIRTHS)	Corr(EXIT, DEATHS)	SAMPLE
CL-NM	0.1792	0.7137	
CL-UCM	0.1913	0.7322	1993:Q2–2013:Q4
ADL-UCM	0.2260	0.7152	

Table 3: Estimated model parameters of univariate Proietti (2006) method for different specifications

REGRESSION MODEL	PARAMETER		ENTRY			EXIT		
ADL(1,1)	Log-Likelihood		-437.6925			-431.91524		
	ϕ		0.8045			0.6200		
	σ^2		$8.9847e^7$			$-1.0435e^8$		
			Value	StDev	t-stat	Value	StDev	t-stat
	Regression Effects	x_1	30716.87	1599.51	19.15	49024.53	1780.91	27.53
		x_2	41.89	18.15	2.31	119.12	20.20	5.90
		x_3	-409716.78	533143.27	-0.77	689839.94	552907.00	1.25
		x_4	769250.14	515659.34	1.49			
Chow-Lin	Log-Likelihood		-439.8661			-431.9442		
	ϕ		0.8015			0.4990		
	σ^2		$-1.0228e^8$			$-1.4756e^8$		
			Value	StDev	t-stat	Value	StDev	t-stat
	Regression Effects	x_1	32623.80	1556.93	20.95	64365.34	2077.61	30.98
		x_2	33.73	18.90	1.78	163.93	24.08	6.81
		x_3	-143156.25	552615.74	-0.26	36358.01	479154.96	0.08
LR-statistic			**4.3471**			1.5835		

Table 4: Stylized Business Cycle Facts

Data Types	Series	Lead	Lag	Sign
Disaggregated (Univariate UCM)	Entry-BK		3	+
	Entry–HP		3	+
	Exit–BK	6		-
	Exit–HP	6		-
Disaggregated (Combined UCM)	AVEntry-BK		3	+
	AVEntry–HP		3	+
	AVExit–BK	6		-
	AVExit–HP	7		-
BLS	Births-BK		0–1	+
	Births–HP		0–1	+
	Deaths–BK	4		-
	Deaths–HP	3		-

Table 5: VAR M1 estimates

Parameter	Summary Statistics
T	146
p	20
Log-Lik	-1113.8585
R^2(LR)	0.7875

Variable	ΔLRGDP			ΔLEntry			ΔLExit			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.4123	0.0850	0.0000	-0.4105	0.4884	0.0331	0.7237	0.3306	0.0302	0.7237	0.3306	0.0302
ΔLRGDP(1)	0.3884	0.1172	0.0012	0.5987	0.2620	0.0238	-1.2767	0.6733	0.0600	-0.0261	0.4558	0.9544
ΔLEntry(1)	-0.01899	0.0227	0.0405	0.8166	0.0508	0.0000	-0.0652	0.1307	0.6185	-0.0672	0.0885	0.4485
ΔLExit(1)	0.0374	0.0147	0.0123	0.1027	0.0329	0.0022	0.1176	0.0847	0.1674	0.1460	0.0573	0.0120
ΔLTFP(1)	-0.0126	0.0850	0.6968	-0.0077	0.0726	0.9156	0.0250	0.1865	0.8936	0.1220	0.1263	0.3355
σ		0.7127			1.5936			4.0948			2.7721	
RSS		71.6380			358.1116			2364.2790			1083.594	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degree of freedom)	Statistic (value)	p-value
Portmantau	Resid. Correlation	$\chi^2(176)$	360.38	0.0000**
Godfrey	Autocorrelation	F(32,481)	4.9947	0.0000**
Doornik-Hansen	Normality	$\chi^2(8)$	108.32	0.0000**
White	Heteroskedasticity	F(32,495)	2.7309	0.0000**
RESET	Correct Specification	F(32,481)	1.6192	0.0190*

Table 6: VAR M2 estimates

Parameter	Summary Statistics
T	146
p	20
Log-Lik	1972.1950
R^2(LR)	0.9295

Variable	ΔLRGDP			ΔLAVEntry			ΔLAVExit			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.0040	0.0008	0.0000	0.7139	0.3455	0.0406	0.0017	0.0032	0.5844	0.0003	0.0003	0.2724
ΔLRGDP(1)	-0.0106	0.0816	0.8959	0.6903	0.0619	0.0000	0.0766	0.4089	0.8517	-0.0101	0.0244	0.6798
ΔLAVEntry(1)	-0.0306	0.0229	0.1832	0.0456	0.0317	0.1532	-0.0405	0.0963	0.6743	-0.0140	0.0074	0.0600
ΔLAVExit(1)	-0.0210	0.1913	0.0000	-0.0110	0.4948	0.9823	0.5943	0.0490	0.0000	9.8e-05	0.0075	0.9897
ΔLTFP(1)	1.2794	0.1913	0.0000	-0.0048	0.0029	0.1096	-0.6550	0.7138	0.3603	0.8650	0.0606	0.0000
σ		0.0063			0.0196			0.0245			0.0018	
RSS		0.0056			0.0543			0.0848			0.0004	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degree of freedom)	Statistic (value)	p-value
Portmantau	Resid. Correlation	$\chi^2(176)$	492.57	0.0000**
Godfrey	Autocorrelation	F(80,467)	6.5140	0.0000**
Doornik-Hansen	Normality	$\chi^2(8)$	107.09	0.0000**
White	Heteroskedasticity	F(32,495)	2.9213	0.0000**
RESET	Correct Specification	F(32,481)	2.0567	0.0007**

Table 7: VAR Models description

Model	Sample	Y_t (in Δlog)
M1	1977:Q1–2013:Q4	[RGDP Entry Exit TFP]'
M2		[RGDP AVEntry AVExit TFP]'
M3	1992:Q3–2013:Q4	[RGDP OPENINGS CLOSINGS TFP]'
M5		[RGDP Entry Exit TFP]'
M6		[RGDP AVEntry AVExit TFP]'
M4	1993:Q3–2013:Q4	[RGDP BIRTHS DEATHS TFP]'
M7	1977:Q1–1993:Q2	[RGDP NBI Exit TFP]'