ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Firms' Dynamics and Business Cycle: New Disaggregated Data (updated version)

Lorenza Rossi
(Università di Pavia)

Emilio Zanetti Chini
(Università di Pavia)

 # 141 (06-17)

Via San Felice, 5
I-27100 Pavia
http://epmq.unipv.eu/site/home.html

June 2017

Firms' Dynamics and Business Cycle: New Disaggregated Data

LORENZA ROSSI AND EMILIO ZANETTI CHINI*

University of Pavia
Department in Economics and Management
Via San Felice 5 - 27100 Pavia (ITALY)
e-mail: emilio.zanettichini@unipv.it

FIRST VERSION: May 2016
THIS VERSION: June 2017

Abstract

We provide stylized facts on firms dynamics by disaggregating U.S. yearly data from 1977 to 2013. To this aim, we use an unobserved component-based method, encompassing several classical regression-based techniques currently in use. Our 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. A structural econometric analysis supports the findings of the most recent theoretical literature on firms dynamics. Standard macroeconometric models estimated from our data outperform their equivalents estimated using the existing series.

JEL Codes: C13, C32, C40, E30, E32.

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

*Corresponding Author.

1 Introduction

Firms are one of the pillars of the economy. Their dynamics is an important indicator of the status of the economic activity. As a consequence, having a precise view of the entry and exit of firms from the market is fundamental in order to detect the business cycle phase.

This appears so elementary that any investigator would be tempted to think that that a myriad of datasets are available to conduct empirical analysis on this topic. Instead, this is only partially true. Namely, the microeconomic dataset currently published at yearly frequency by Census Boureau constitutes the only long-span dataset for empirical studies on the argument; on the other side, the quarterly time series on firm's dynamics - published by the Boureau of Labor Statistics, start in 1992:Q3. This means that the only time series nowadays available for the estimation of macroeconomic models faces with limited – and limiting – frequency.

The scarcity of macroeconomic data at higher frequency motivates this paper. In this respect, our contribution is twofold. First, we improve the availability and quality of the US data by disaggregating the yearly series provided by the Census Bureau. Despite its rich variety of indicators at different level of aggregation, the yearly frequency limits the number of observations to use for econometric analysis. On the other hand, the existing quarterly series available from other sources start in Nineties. Consequently, they do not allow a comprehensive long-span business cycle analysis as required by modern macroeconometric literature. This is exactly our second contribution. In particular, we evaluate the performance of our series either via univariate diagnostics and comparison of our data with the analogue ones by Boureau of Labor Statistics. All these series - the disaggregate ones and the equivalent quarterly ones - are here generically labelled as "ENTRY" and "EXIT", while, in the rest of the paper, such labels will be substituted to correctly identify each single series used.

In the first part of this paper, we consider two different types disaggregation techniques: (i) the one proposed by Chow and Lin (1971), which is still widely used by Statistical Institutes; and (ii) two models based on unobserved components methods (UCM), originally proposed by Proietti (2006). The first technique is a simple OLS estimation of an autoregression with AR(1) errors. The two models based on UCM are instead represented by a unifying state space system and their estimation rely on an Augmented Kalman Filter (AKF). Thanks to this, the UCM allows generality and flexibility besides maintaining statistical robustness.

The series disaggregated by UCM are considerably more accurate and credible of the ones resulting by applying naive Chow-Lin model. Furthermore, the selecting procedure of the two UC models suggests that series of EXIT should be fitted by an AutoRegressive model of order 1 – AR(1), henceforth – while the series of ENTRY is better represented by an UC representation of an Autoregressive Distributed Lag model of same oder – ADL(1), henceforth. Notice that, both the UC models rely on univariate estimations and thus consider a single regressor variable, here represented by the industrial production. The use of a single regressor might be sub-optimal in a disaggregation exercise. This the main limitation of this class of univariate techniques. We provide a first-attempt to circumvent this problem by re-applying extensively the new (univariate) UCM on all the 134 macroeconomic indicators of MD-FRED dataset. The disaggregate variables of ENTRY and EXIT are then computed as a simple average of the resulting 134 univariate UC 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 very similar to the ones derived from the single indicator.

After a comprehensive discussion of data and methods, we use these new disaggregated series for business cycle analysis. We extract the trend and cycle components via standard first difference operator, the Hodrick-Prescott (HP) and Baxter-King (BK) filters. Once the series have been filtered, we classify them 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. A simple cross-correlation analysis allows us to classify the ENTRY series as a lag and procyclical indicator of the business cycle. On the contrary, the EXIT series is a countercyclical leading indicator.

Finally, we run a structural analysis and show the responses of ENTRY, EXIT - jointly with real GDP - to a productivity shock, as suggested by Gali (1999). The results briefly summarized above seem compatible with recent macroeconomic literature on endogenous firms dynamics[1]. Indeed, we find that a productivity 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, our analysis outperforms the same structural econometric models estimated using quarterly BLS series, commonly used in the literature.

Our paper is related to the literature on DSGE models of firms dynamics and its effects on business cycle. At the current state, the theoretical contributions on firms' dynamics are mainly focused on firms's entry[2]. Rossi (2015) and Hamano and Zanetti (2015) cover this theoretical gap in the literature by showing that firms' exit represents an even stronger propagation mechanism of the business cycle than that of firms' entry. Unfortunately, these efforts to give a theoretical explanation of business dynamics face with the data limitations previously introduced. Overcoming this problem is our aim.

The rest of this paper is organized as follows. Section 2 deals with the problem of data availability on firms 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 5 discusses the relevance of their business cycle movements for Macroeconomics by estimating a set of Structural VAR models. Finally, Section 6 concludes. The associated Supplement complements the results of the paper by providing mathematical details on statistical methods here adopted and further results.

2 Data

The only official long-span dataset for U.S. 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 derived measures at yearly frequency from 1977 up to 2014 (current release). In turn, BDS data summarize the confidential data of Longitudinal Business Databases, a census of business establishments and firms covering all industries in all US[3].

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 employment 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; 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. In the course of this paper we will refer to these series with the "OPENINGS" and "CLOSINGS" labels.

   Alternatively, it is also possible to use the two series of establishment birth and death for total private sector, (BIRTHS" and "DEATH" - or B and D - respectively, henceforth) available from the same source from 1993:Q2. These series are direct observations of the number of firms/establishments.

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 low of motion of firms' entry without any empirical counterfactual. Perhaps more astonishing, this series is the basis of a recent strand of literature[4] despite the fact that the exact definition of incorporation used is not available.

Ergo, firms' dynamics is not observed and commonly measured by a proxy - the employment level - quite imperfect. This is evident from a look at the cross-correlation function of these two series with the lags of series of real gross domestic product – RGDP, henceforth[5] – shown in Figure 1: the estimated bars are largely below the critical values of ± 0.20, corresponding to the blue bars. Our contribution is precisely in the production of new time series on firm's dynamics capable to overcome this measurement error problem.

Finally, as the next Section exposes, the disaggregation of a low-frequency time series is strictly related to the choice of an high frequency series to use as benchmark. What time series should be adopted? In principle, any possible macroeconomic time series generically related to the estimation of GDP aggregate. Hence, we adopt the FRED-MD, a new macroeconomic database of 134 indicators of all sectors of the U.S. economy, recently published by Federal Reserve Bank of St. Louis and extensively described in McCracken and Ng (2016)[6]. Namely, we regress the yearly ENTRY (EXIT) series on all these variables in quarterly frequency by in order to obtain a new quarterly series of the regressand. Anyway, as the next sections will document, this solution conveys a time series so similar to the one obtained via univariate regression that further methodological investigation seems us purely academic.

3 Temporal Disaggregation

In a temporal disaggregation exercise, measurements of some variable are available only 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/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 can be solved via *interpolation* or via *distribution*. The former consists in estimating 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 describe the Chow and Lin (1971) method. This relies on a simple OLS regression of an AR(1) process with the observed indicators as exogenous regressors, which distribution function is assumed to be known. For this reason we label this the Chow-Lin Naive Method (CL-NM). We then discuss the Proietti (2006) UCM. This consists of two models: (i) the first one is a UC representation of the Chow-Lin regression model (CL-UCM); (ii) the second one is a more general autoregressive distributed lag (ADL-UCM) model. Both of them are estimated by AKF[7]. Finally, in the last part of this Section, we propose to apply the univariate UCM to MD-FRED dataset, where we repeat the same UCM for all the indicators therein contained. The disaggregated ENTRY (EXIT) will be then computed as a simple average of all quarterly series deriving from disaggregation of each single indicator. Thus, we label this Combination Method.

In what follows, y_t denotes a realization of a (univariate) time series observed at $t = 1 \ldots, t, \ldots, T$; multivariate time series are denoted in bold and matrices in capital.

3.1 The CL-NM

Let assume, for easy of explanation, that we are interested in univariate disaggregation, that is to disaggregate a single time series using a single indicator variable. This is a simple linear OLS regression of the observed $T \times 1$-dimentional process y_t on the vector of the same dimensions of the indicator variable x_t multiplied by the $T \times sT$ disaggregation matrix D, i.e.

$$\begin{aligned} D_s' y_t &= D_s' x_t' \beta + D_s' u_t, \quad u_t \sim N\left(0, D_s V D_s'\right) \\ u_t &= \phi u_{t-1} + \epsilon_t, \;\; |\phi| < 1, \;\; \epsilon_t \sim NID(0, \sigma^2), \end{aligned} \tag{1}$$

where D_s is a block-diagonal matrix of form

$$D_s \doteq \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

for the case in which $s = 4$, that is from yearly to quarterly frequency.

In order to make (1) equivalent to the CL-NM, three noticeable assumptions on x_t must be done:

CL 1. *x_t is strictly exogenous;*

CL 2. *x_t free of measurement errors;*

CL 3. *x_t is cointegrated with y_t.*

Remark 1. The univariate framework here adopted does not affect the generality of these assumptions, since they hold for all N possible regressors, that is with $\boldsymbol{X_t} = [x_{1,t}, \ldots, x_{N,t}]$ in place of x_t.

3.2 The Unobserved Component Method

Assumptions CL1 – CL3 are quite restrictive. In particular, the contribution by Harvey (1989) stresses that CL2 is limiting in many common empirical applications. Here we discuss the way to overcome them. To this aim we rely on the following state space representation for the general model for temporal disaggregation:

$$\begin{cases} y_t = z'\alpha_t + x_t'\beta, \\ \alpha_t = \alpha_{t-1} + W_t'\beta + H\epsilon_t \\ \alpha_1 = a_1 + W_1\beta + H\epsilon_1 \end{cases} \tag{2}$$

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, corresponding to the indicators, that enter respectively the measurement and the transition equations and zero elements corresponding to effects that are absent from one or the other equations. 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 might differ from H), and regression effects, $W_1\beta$.

For what follows the following assumptions are invoked:

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

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

Assumption H1 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

H2 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 these assumptions are just for convenience and can easy be weaken[8]. Under this general framework, the model (1), properly arranged, can be assumed as particular case of (2).

Case 1 (CL-UCM model)**.** When α_t is a scalar, $z = 1$, $T = \phi$ are $H = 1$, the system (2) degenerates into a linear regression model with AR(1) errors:

$$y_t = z'\alpha_t + x_t'\beta, \quad \alpha_t = \phi\alpha_{t-1} + \epsilon_t, \quad \epsilon_t \sim NID(0, \sigma^2) \tag{3}$$

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

Remark 2. 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 (2) degenerates into an ADL model:

$$y_t = \phi y_{t-1} + m + gt + x'\beta_0 + x_{t-1}'\beta_1 + \epsilon_t, \tag{4}$$

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) for further details.

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

Remark 3. The ADL-UCM collapses to the CL-UCM if and only if:

$$\beta_1 = \phi\beta_0. \tag{5}$$

Thus, (4) can be rewritten as:

$$y_t = x_t'\beta_0 + \alpha_t, \quad \alpha_t = \phi\alpha_{t-1} + \epsilon_t \tag{6}$$

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

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

Proposition 1. Let define the cumulator variable as:

$$y_t^c \doteq \psi_t y_{t-1}^c + \psi_{t-1}, \quad \psi_t^c \doteq \begin{cases} 0 \text{ if } t = s(\tau-1)+1, \text{ with } \tau = 1, \dots, [n/s] \\ 1 \text{ otherwise} \end{cases} \tag{7}$$

Then the State Space System

$$\begin{cases} y_t = z^{*'}\alpha_t^* \\ \alpha_t^* = T_t\alpha_{t-1}^* + W_t^*\beta + H^*\epsilon_t, \\ \alpha_1^* = a_1^* + W_1^*\beta + H_1^*\epsilon_1, \qquad \epsilon_t \sim NID(0,\sigma^2), \quad \beta \sim N(b, \sigma^2 V), \end{cases} \tag{8}$$

where:

$$z^* = \begin{bmatrix} 0' \\ 1 \end{bmatrix}, \; T_t^* = \begin{bmatrix} T & 0 \\ z'T & \psi \end{bmatrix}, \; W_t^* = \begin{bmatrix} W_t \\ z'W_t + x_t' \end{bmatrix}, \; H_t^* = \begin{bmatrix} H_t \\ z'H_t + x_t' \end{bmatrix}, \tag{9}$$

$$\alpha_1^* = \begin{bmatrix} a_1 \\ z'a_1 \end{bmatrix}, \; W_1^* = \begin{bmatrix} W_1 \\ z'W_1 + x_1' \end{bmatrix}, \; H_1^* = \begin{bmatrix} H_t \\ z'H_1 + x_1' \end{bmatrix} \tag{10}$$

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 (7), substituting the transition equation and re-writing:

$$y_t^c = \psi_t y_{t-1}^c + \psi_{t-1} + z'T\alpha_{t-1} + (z'W_t + x_t\beta) + z'H\eta_t, \tag{11}$$

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

3.2.2 Estimation

Estimation of system (8) is done by Maximum Likelihood. Depending on the presence of diffuse elements in matrix W or not, two estimators are possible:

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

$$\hat{\beta} = -S_{n+1}^{-1}s_{n+1} \quad ,Var(\hat{\beta}) = S_{n+1}^{-1}, \quad \hat{\sigma}^2 = \frac{q_{n+1} - s_{n+1}'S_{n+1}^{-1}s_{n+1}}{[n/s]}, \tag{12}$$

with profile Log-Likelihood:

$$\mathcal{L}_{\mathcal{F}} = -0.5[d_{n+1} + [n/s](\ln \hat{\sigma}^2) + (\ln \hat{\sigma}^2 + \ln 2\pi + 1)]. \tag{13}$$

Diffuse: $\hat{\beta}$ 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 Log-Likelihood:

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

In 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 4. 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 (2), 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. The annexed Supplement provides mathematical details.

3.3 Combination Method

The UCM disaggregation previously described holds for a single time series, that is, In terms of system (8), we are assuming x_t is $[T \times 1]$ vector. Such an assumption is hard to justify in many 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 and run the UCM for disaggregation machinery to all the N elements of X_t. In this way, one can use informations from all the sectors of the economy. Consequently, we produce N differently disaggregated processes collected in $[T \times N]$ matrix $\mathbf{\hat{Y}_t} = [\mathbf{\hat{y}_1}, \ldots, \mathbf{\hat{y}_n}, \ldots, \mathbf{\hat{y}_N}]$, with y_n representing the $[T \times 1]$ vector of the process disaggregated via system (2) 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 of (8) and (12)–(13) or (12)–(14)

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

$$\hat{y}_t^{\,C} \doteq \frac{1}{N}\sum_{n=1}^{N} y^s_{t,n}, \tag{15}$$

with $y^s_{t,n}$ denoting the time series at time t disaggregated according to the univariate UCM using the n-th indicator. The equation (15) conveys the series labeled $AVEntry$ and $AVExit$ – to underline the fact that they are an average of many single processes.

4 Business cycle analysis

A first graphical inspection of the series outcoming from all the three models considered in previous section is shown in Figure 2. In particular, we notice that the series obtained by CL-NM is completely different to the others at the end of sample (for example in 2013 the annual average level of EXIT is 170,000 in the CL-NM against 188,000 for the equivalent series measured by the CL-UCM). Another difference between the CL-NM and UCM, emerges when looking at the correlations between our disaggregated series of ENTRY (using different techniques) and the quarterly data of establishments OPENINGs (BIRTH) and, in a similar fashion, between disaggregated series of EXIT and the series of CLOSING (DEATH). These are reported in Table 1: there is a large spread between correlations of ENTRY and EXIT measures (0.23 and 0.70 in mean, respectively); moreover, the correlations of our series obtained CL-NM and the OPENING/CLOSING proxies are 40% lower than the UCM equivalent ones. In addition to this, the Johansen (1991) test reject strongly the null hypothesis of one cointegrating relation between the series of ENTRY and IP – properly aggregated[9]. Being full cointegration of indicators a requirement for applying the UCM, this finding ends our investigation on the CL-NM, see Remark 2 in Section 3.2.

The series estimated via CL-UCM specification are nicely smooth. Interestingly, this "smoothness" is lower if ADL specification is selected. This is immediately visible in periods of recession. Let consider the example of 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 the ADL-UCM one ranges from 181,500 to 167,000 – that is, the level is more than the double, in absolute value, under recession. The quasi-noisy behaviour of ADL specification characterizes all the span of the series.

Thus, the difference in the estimates deriving from the two specifications of the UCM is not negligible. What specification of UCM should be chosen? This is a typical selection problem that can be solved via simple Likelihood-Ratio (LR) test. According to its high LR-statistic (4.3) reported in Table 2, the model chosen for ENTRY is the ADL-UCM. On the contrary, the LR-statistics of EXIT (1.5) is not able to reject the CL-UCM specification.

The LR-test is a criterion to select the model specification to use when combining the data disaggregated by the same UCM from using all the 134 indicators of FRED-MD. This leads to new time series qualitatively very similar to the ones derived from a single indicator previously analyzed.

We are ready to use our disaggregated series of ENTRY and EXIT for business cycle analysis, which is graphed in Figure 5. All the series have the same path in both trend and cycle components, and just in the trend some difference is visible between the two filters used; the cycle component of univariate and Combination UCM deliver is substantially coincident. If instead we repeat the same graphical check on BLS data, the evidence gives a different picture: as shown in Figure 6, the BK filter conveys a cycle considerably more smooth than the HP one. On the other hand, the HP-filtered OPENINGS series has a wildly noisy trend and a short cycle (very similar to the one of BIRTHS), while the CLOSINGS and DEATHS are more nicely smooth and have a dynamics different in terms of peaks and troughs; it is interesting to notice that BIRTHS (DEATHS) are more (almost) than the double of OPENINGS (CLOSINGS) in the 2007–2009 recession. These differences in the cycles can be better appreciated if comparing the series from all different sources;

this is done in Figure 7: our series anticipate the recession phase in ENTRY, albeit BLS data are "leading" in EXIT measures. At the light to these findings, the nature of the indicators we are working with is different
Thus, we need to definitively classify our disaggregated series as leading or lagging indicators of the economic activity. To this aim, we compute the maximum absolute value cross-correlations between the cycle of disaggregated series and that of the RGDP. 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 series and 8 quarters for the EXIT series[10]. According to Table 3 we find the following results:

- the disaggregated ENTRY series from BDS is generally a lag and pro-cyclical indicator of the business cycle with maximum absolute cross-correlation at the lag 3 for the BK and the HP filter and at the lag 8 for the linear difference filter, which constitutes the only difference with Combination UCM series.

- On the contrary, the Combined UCM EXIT series is negatively correlated with RGDP and leads the cycle with a maximum cross-correlation at lag 6 for the BK and the HP filter and at lag 7 for the linear difference filter. Moreover, the equivalent Univariate UCM series are considerably different: the BK-cycle is counter-cyclical and leading; instead, the opposite holds for the HP cycle (pro-cyclical with two years of lagging).

- The contemporaneous correlations with RGDP are positive for ENTRY and negative for EXIT. However, they are not statistically significant. Differently, both of the two BLS proxies are generally pro-cyclical. Namely, OPENINGS and BIRTHS are coincident indicators, while CLOSINGS (DEATHS) are generally lagging (leading).

5 Structural 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 adopt a Structural Vector Auto-Regression (SVAR) model, which will be briefly introduced in the next Subsection 5.1; see Lütkepohl (2005) for details on estimation, modelling and forecasting methods in this framework. Then, in Subsection 5.2, we consider five specifications of that model for different combination of Entry and Exit proxies. Table 4 gives a complete description of the samples and system of each model, which is labeled as M1 – M5. An important *caveat* is that the order of the VAR system is 1 in all the cases for allowing a comparison with the literature and the optimal properties of the resulting dynamic multipliers.

5.1 Prototypical VAR models

Let consider the M1 as an example. In this case, $\boldsymbol{y_t} = [RGDP_t, ENTRY_t, EXIT_t, TFP_t]'$ is a N vector of observable time series of length T with $T = 144$ at the full sample and having the reduced-form:

$$\boldsymbol{y_t} = \boldsymbol{A_1 y_{t-1}} + \boldsymbol{u_t}, \tag{16}$$

where $\boldsymbol{A_j}$, $(j = 1 \dots p)$ are $(N \times N)$ coefficient matrices and $\boldsymbol{u_t} = [u_{RGDP,t}, u_{ENTRY,t}, u_{EXIT,t}, u_{TFP,t}]'$ is a white noise vector of time series with $\boldsymbol{u_t} \sim (\boldsymbol{0}, \boldsymbol{\Sigma_u})$. The eq. (16) has the following Wold representation:

$$\begin{aligned} \boldsymbol{y_t} &= \boldsymbol{\Phi_1 u_{t-1}} + \boldsymbol{\Phi_2 u_{t-2}} + \dots, \\ \boldsymbol{\Phi_j} &= \sum_{j=1}^{k} \boldsymbol{\Phi_{k-j} A_j}, \\ \boldsymbol{\Phi_0} &= \boldsymbol{I_N}, \end{aligned} \tag{17}$$

where k=1, 2, ..., and $\boldsymbol{\Phi_j}$ is the Impulse Response Functions (IRFs).

In order to orthogonalize the shocks, we adopt two strategies:

1. Zero Short-Run Restrictions via simple Cholesky decomposition of covariance matrix $\boldsymbol{\Sigma_u} = \boldsymbol{PP'}$, where $\boldsymbol{P}$ is a lower triangular matrix with positive elements in the main diagonal. In our exercise, we assume that the shock due to technology are the most exgoenous. Thus we have the following ordering: (i) TFP_t, (ii) $RGDP_t$, (iii) $Entry_t$ and (iv) $Exit_t$.

2. Zero Long-Run Restrictions via decomposition á la Blanchard and Quah (1989). In this case, the error vector is reparametrized as follows:

$$\boldsymbol{A}u_t = \boldsymbol{B\epsilon_t}, \quad \boldsymbol{\epsilon_t} \sim (\boldsymbol{0}, \boldsymbol{I_k}), \tag{18}$$

   where $\boldsymbol{u_t} = \boldsymbol{A^{-1}B}\epsilon_t$ and, consequently, $\boldsymbol{\Sigma_u} = \boldsymbol{A^{-1}BB'A^{-1'}}$. From eq. (18) – also known as "AB" model as defined in Amisano and Giannini (1997) – one can derive the Total Impact Matrix:

$$\boldsymbol{\Xi} = \sum_{i=0}^{\infty} \boldsymbol{\Theta_i} = (\boldsymbol{I_k} - \boldsymbol{A})^{-1}\boldsymbol{B}, \tag{19}$$

   whose elements are set to zero where there is assumed that the shocks do not have long-run effect on other variables. In M1, for example, the total impact matrix is:

$$\boldsymbol{\Xi} = \begin{pmatrix} \theta_{11} & 0 & \theta_{13} & \theta_{14} \\ 0 & \theta_{22} & \theta_{23} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} \end{pmatrix}. \tag{20}$$

In what follows, all shocks are assumed being one standard deviation TFP shock[11]. All series in all systems are in logarithmic differences.

5.2 Results

Tables 5 – 9 report the estimates of M1 – M5[12]. It is immediate to notice that (AVE)ENTRY/ENTRY are non-normal. On the other side, the gaussian hypothesis holds for BLS proxies (BIRTHs and DEATHs in M3) and – albeit very weakly – for our disaggregated variable if considering a short sample (M4). All the models present additional serial correlation. However, augmenting the lag order of the VAR system is critical for a matter of policy implications (IRFs become radically different and "gross"). On a different side, in short samples like our ones, the heteroskedasticity is symptomatic of a possible misspecification of the model as confirmed by the results of RESET. In many applications these findings (non-normality, heteroskedasticity and misspecification) are symptomatic of a nonlinear behaviour of the series. Nevertheless, they do not preclude to do a structural analysis. Thus, the resulting IRFs are plotted in Figures 8 – 12. Three findings are noticeable:

- As expected by theory, the RGDP increases in response to the shock, ENTRY is pro-cyclical and persistent. On the contrary, EXIT is countercyclical in the medium run, while it reacts positively on impact, even though this positive impact response might be not statistically significant according to the confidence bands. Hence, the statistical findings of Table 3 are confirmed.

- In general, the responses of RGDP are almost immediate (between 1 and 3 quarters) and of TFP just quite less (between 3 and 5 quarters). On the other hand, very noticeable variations occur in the responses of ENTRY and EXIT. According to the theory, one should expect an immediately positive (negative) response for ENTRY (EXIT); this is holds only partially for M1 (where univariate UCM series are employed), while M2 (Combined UCM) is almost "theory consistent". The use of a different dataset, as for the case of M4, helps to augment the clearness of the same IRF; however, this has the cost to loose a considerable number of observations.

- What happen if considering our disaggregated data to the same time-span of BLS? M4 and M5 show clearly that the responses of ENTRY and EXIT are possibly more extreme but at cost of an overshooting effect. This is particularly evident in the EXIT series of M6.

6 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 an unobserved component-based temporal disaggregation method to data on entry and exit of firms at establishment level.

Our new time series of entry and exit of firms at establishment level are feasible proxies of Business Cycle. In particular, EXIT is a leading and countercyclical indicator, while ENTRY is lagging and procyclical. Moreover, our new quarterly BDS series on firms' dynamics shows that a TFP shock is associated to a negative and persistent response of EXIT and a positive and persistent response of ENTRY.

The availability of new macroeconomic data allows us to gain additional information to use for our estimates. No matter of this, 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 combination strategy. We think that such a result can be explained from the simple fact that many of the variables in the FRED-MD 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. This could be an interesting development in light of new advances in variable selection algorithms as in Bai and Ng (2009).

Finally, the diagnostic checks in our SVAR models lead to think that our disaggregated series could be affected by nonlinearity. This is partially confirmed by Zanetti Chini (2017), and could be due to the fact that we used a nonlinear variable (the industrial production) as indicator. Thus, we are confident that a properly setted nonlinear structure could be useful to implement descriptive properties of the process and to forecasting aims.

Notes

1 See, *inter alia* Bilbiie *et al.* (2012); Rossi (2015); Lewis (2009); Etro and Colciago (2010); Colciago and Rossi (2015).

2 In their seminal article, Bilbiie *et al.* (2012) - BGM, henceforth - 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. Ergo, they are not able to disentangle the role of firms exit with respect to firms' entry.

3 For more details, see the Census web page, where data are available. When this paper was initialized, we adopted the 2015 release the time span of the data was up to 2013. We maintain this release to avoid problems of variation in the estimates that often characterizes official datasets, as in our case.

4 See Bergin and Corsetti (2008); Lewis and Poilly (2012); Bergin *et al.* (2014); Lewis (2013) *inter alia.*

5 Data source: FRED, Federal Reserve Bank of St. Louis.

6 The FRED-MD is at monthly frequency and, at the time of the settlement of our investigation, it was the only available online. Since in our application the focus is in quarterly data, we aggregated the original data available from the release 2015-04. At the present date, the quarterly version of the dataset, FRED-QD, is ready for the use. All the monthly releases and the quarterly version can be downloaded at the Michael McCraken's web-site.

7 Both of 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.

8 In particular, β_t could be assumed to vary over time, see Durbin and Koopman (2012) for more complex specification of the latent component model. Anyway, we find that our data do not makes such a complex dynamics more useful than the simpler one here adopted.

9 The Johansen test is a simple likelihood-ratio test on the hypothesis that matrix $Z = [x : y]$ has rank at least zero; if rejected, it is possible to continue the investigation for higher ranks up to the hypothesis of rank of $n-1$, for n denoting the number of variable in the VAR system; here, we assumed a bivariate VAR so the only two alternative hypothesis coincides with no cointegreation vs perfect cointegration. In particular, our application of the test via CATS for RATS software, with 2,500 bootstrap replications, leads to a trace test statistic of 43.766 for the hypothesis of rank 1 and a p-value of less than 1%, and consequently it is strongly rejected. See Juselius (2006) for further methodological details.

10 We enlarged the number of possible lags/leads whenever it was not possible to find a maximum inside this range in order to avoid spurious results. In facts, 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 Altissimo *et al.* (2001, 2010) for a discussion of this problem. In our application, Figure 5 makes us able to notice that the cycle of ENTRY is shorter than the cycle of EXIT. *Ergo*, the choice of a different number of leads/lags for the two disaggregated series.

11 Here, the TFP here used is the series computed by Fernald and Matoba (2009) denominated "Aggregate Productivity". It can be downloaded, jointly with other related measures of productivity, from the website of the Federal Reserve Bank of S.Francisco.

12 Some diagnostic graphical checks on estimated errors, jointly with results for a Long-Run restriction scheme are in Technical Supplement. are available in the Supplement.

Acknowledgments

This paper has been supported by the Project "MACFINROBODS" supported by the European Union. The Authors are grateful to Tommaso Proietti and Ambrogio Cesa Bianchi for having shared their "TDS" and "VARToolbox" packages, respectively, in Ox and MatLab; these packages have been properly modified by the

Authors, which are responsible for any error or imprecision in the calculations. The Authors thanks also Barbara Annicchiarico, Carolina Castagnetti, Cecilia Susanna Jona-Lasinio, Vivien Lewis, Jean Palate, Anthony Savagar, Chris Sims and all participants to ERCIM-CFE2015 in London and 2nd MACFINROBODS in Bruxelles and the Internal Economic Seminar at University of Pavia. Finally, Emilio Zanetti Chini is in debt with the doctors, assistants and nurses of the Department of Emathology of Policlinico "S. Matteo" of Pavia, without whose (free) cares this paper would probably remain unfinished.

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.

Amisano, G. and Giannini, C. (1997). *Topics in Structural VAR Econometrics*. Berlin, 2nd edn.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, **24**.

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.

Blanchard, O. and Quah, D. (1989). The dynamic effect of aggregate demand and supply. *American Economic Review*, **79** (2), 655–673.

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.

Doornik, J. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bullettin of Economics and Statistics*, **70** (1), 927–939.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Models*. Oxford, UK.

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.

GALI, J. (1999). Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? *American Economic Review*, **89**, 249–271.

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.

HENDRY, D. and DOORNIK, J. (1988). *Misspecification tests in Econometrics*. Cambridge.

— and — (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London.

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. *Econometrica*, **59**, 1551–1580.

JUSELIUS, K. (2006). *The Cointegrated VAR Model – Methodology and Applications*. Oxford, UK.

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.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series*. Berlin.

McCracken, M. and Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, **34**, 574–589.

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

Ramsay, J. (1969). Tests for specification errors in classical linear least squares regression analysis. *Journal of Royal Statistical Society, B*, **31**, 350–371.

Rossi, L. (2015). Endogenous Firms' Exit, Inefficient Banks and Business Cycle Dynamics. DEM Working Paper no. 99, University of Pavia.

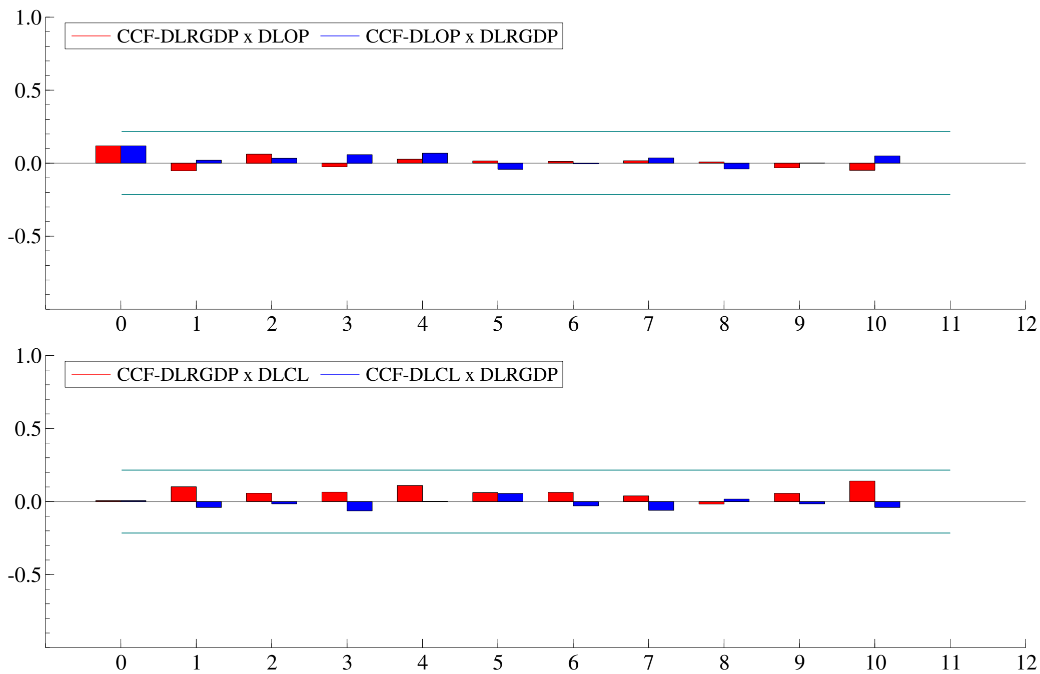
Siemer, M. (2014). Firm Entry and Employment Dynamics in the Great Recession. FEDS Working Paper No. 2014-56.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.

Zanetti Chini, E. (2017). Generalizing Smooth Transition Autoregression. DEM Working Paper No. 138, University of Pavia.

7 Tables and Figures

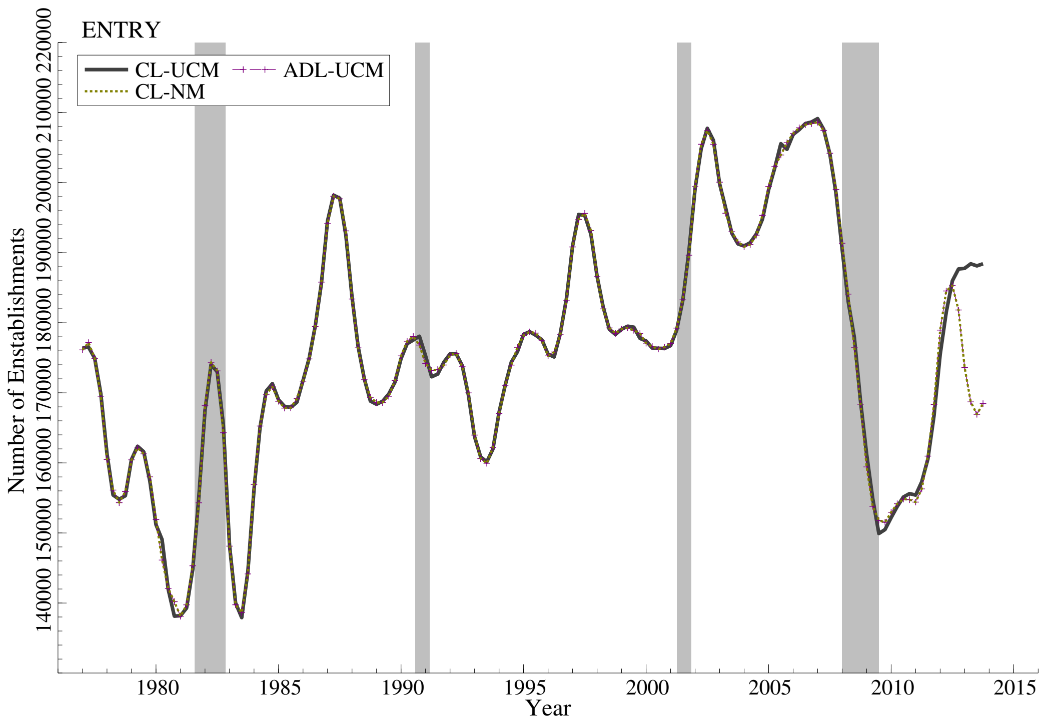
Figure 1: Cross-correlation function of BLS data with RGDP



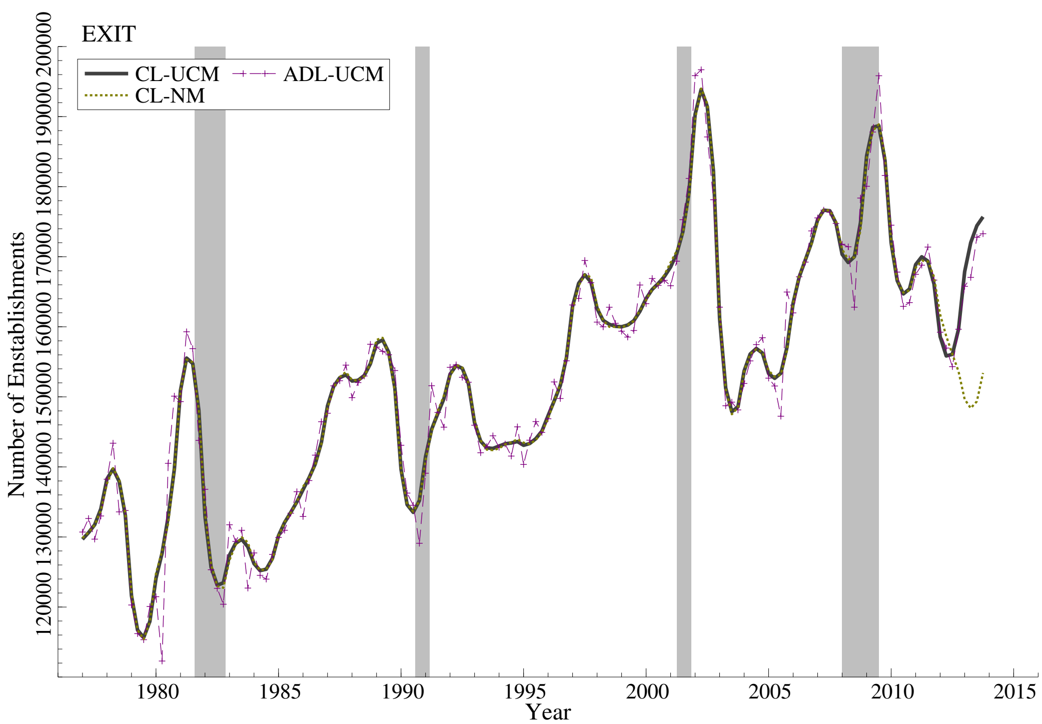
NOTE: This figure plots the cross-correlation function of the BLS data (OPENINGs and CLOSINGs) with RGDP; both the data are in growth rates. Red values indicates leads while blue value indicates lags. The two horizontal lines are the critical values for the significance of the function.

Figure 2: The disaggregated series

(a) ENTRY



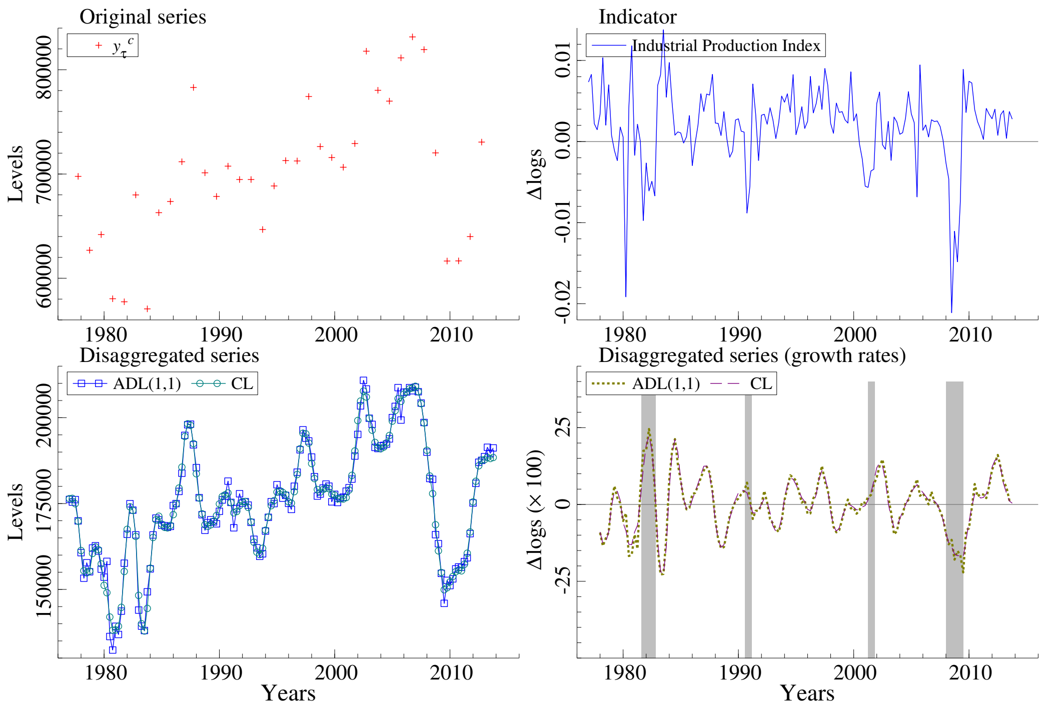
(b) EXIT



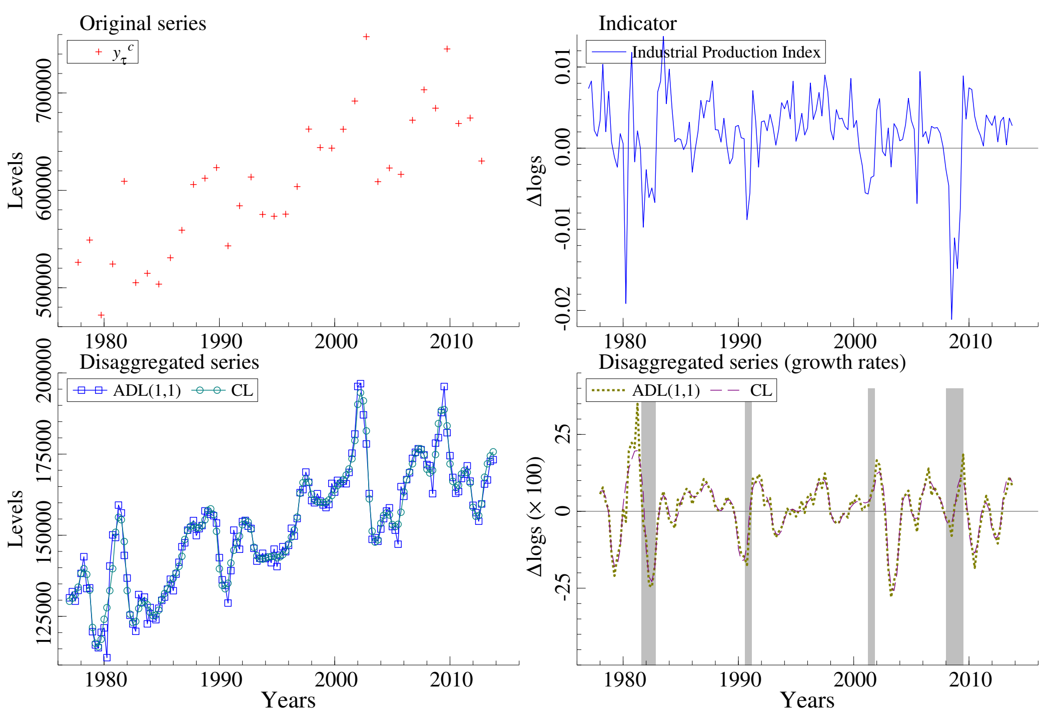
NOTE: This figure plots the values of the time series resulting from the three different temporal disaggregation methods exposed in Section 3. Colored bands correspond to the NBER recessions. Software used: OxMetrics

Figure 3: Comparison of different regression models of univariate UCM

(a) ENTRY

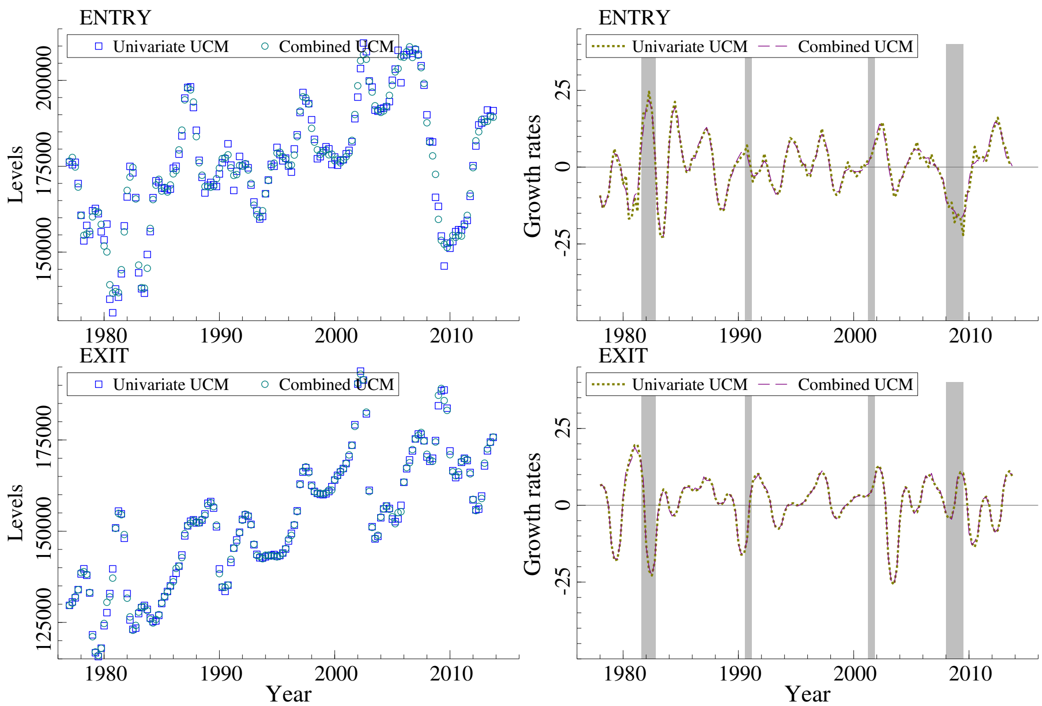


(b) EXIT



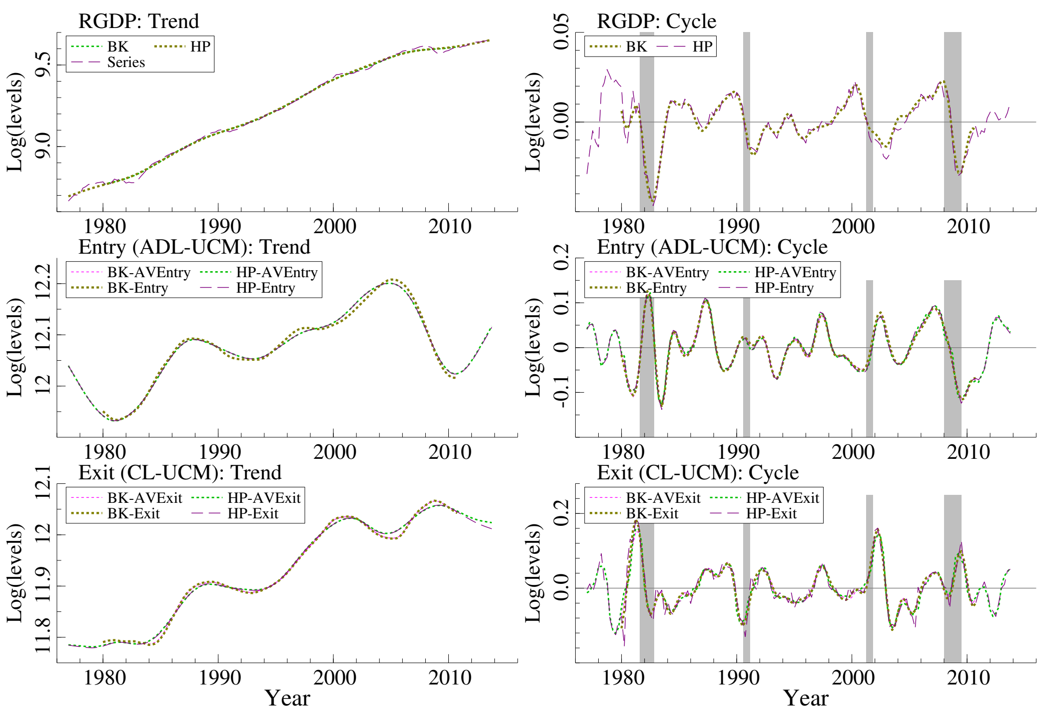
NOTE: This figure plots the ingredients forming the univariate UCM discussed in Section 3.2, and namely: (i) the original yearly-frequency series (upper left panel); (ii) the indicator variable (upper-right); (iii) the resulting disaggregated series according to ADL and CL models of UCM (lower-left); and (iv) the same disaggregated series, but in growth rates (lower-right). Colored bands correspond to the NBER recessions. Software used: OxMetrics

Figure 4: Comparison of univariate and combined UCM.



NOTE: This figure compares the series resulting from the application of univariate disaggregation discussed in Section 3.2 and combined UCM discussed in Section 3.3. The models are selected according to the results of LR-test in Table 2. Upper panels refers to ENTRY and lower panels to EXIT; on the other side, left panels plot series in levels and right panels plot the same series in growth rates. Colored bands correspond to the NBER recessions. Software used: OxMetrics.

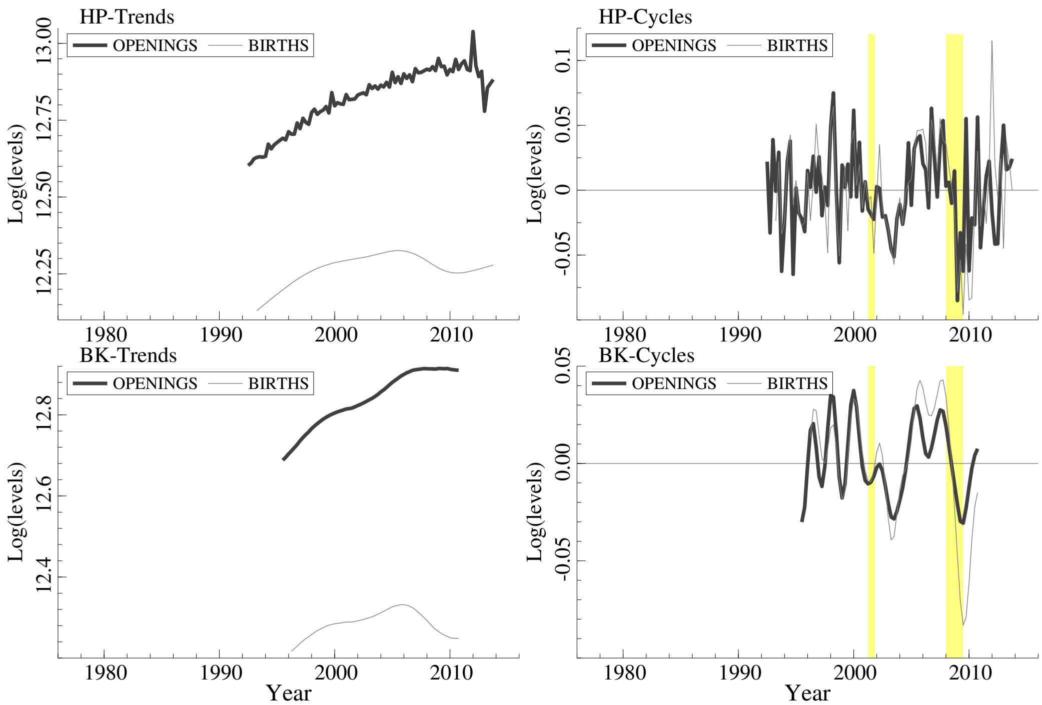
Figure 5: Business cycle analysis of disaggregated series



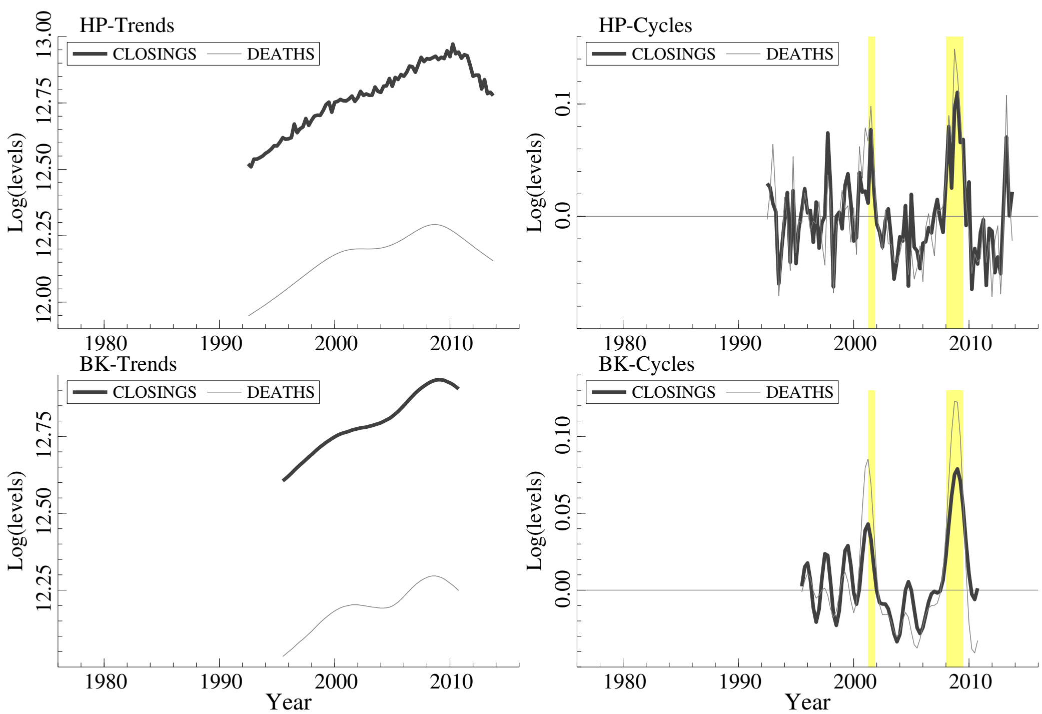
NOTE: This figure shows the business cycle analysis of the RGDP (upper panel) and the ENTRY and EXIT series (central and lower panel, respectively) resulting from the application of univariate disaggregation discussed in Section 3.2 and combined UCM discussed in Section 3.3. Namely, left (right) panels plot the trend (cycle) component extracted from the series in logarithms. The cyclical components extracted by the HP filter assume penalization parameter $\lambda = 1600$ and the ones obtained by BK filter assume parameters $\lambda_0 = 1.5$, $\lambda_1 = 8$ and $K = 12$ (that is, $3 \times s$). Colored bands correspond to the NBER recessions. Software used: OxMetrics.

Figure 6: Business cycle analysis of BLS quarterly series

(a) ENTRY



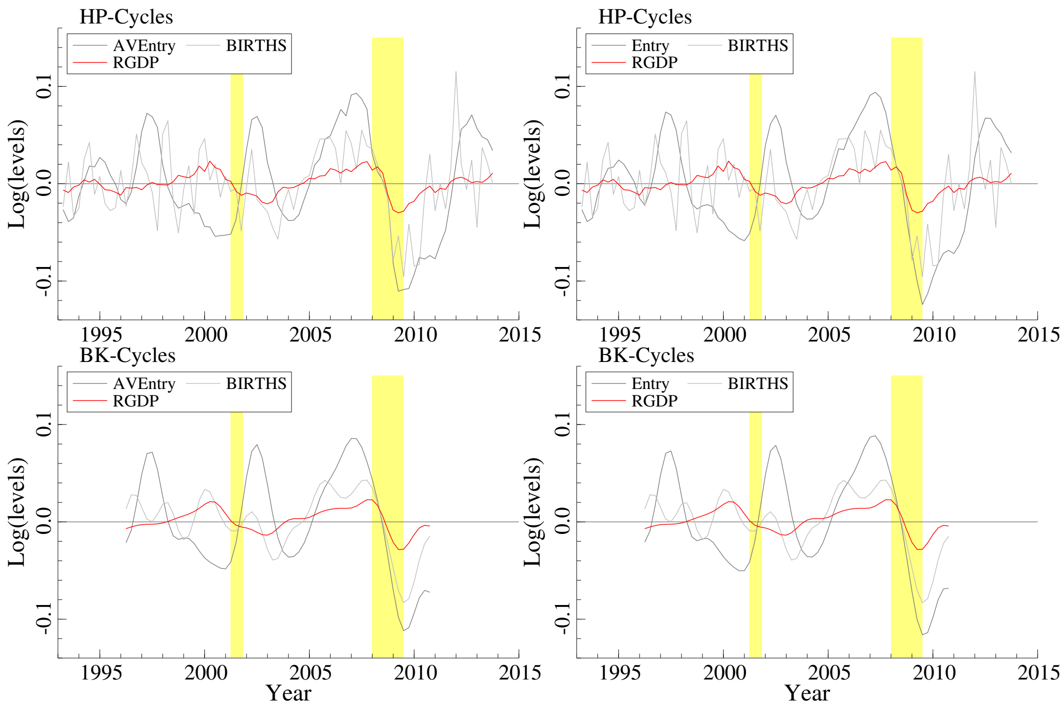
(b) EXIT



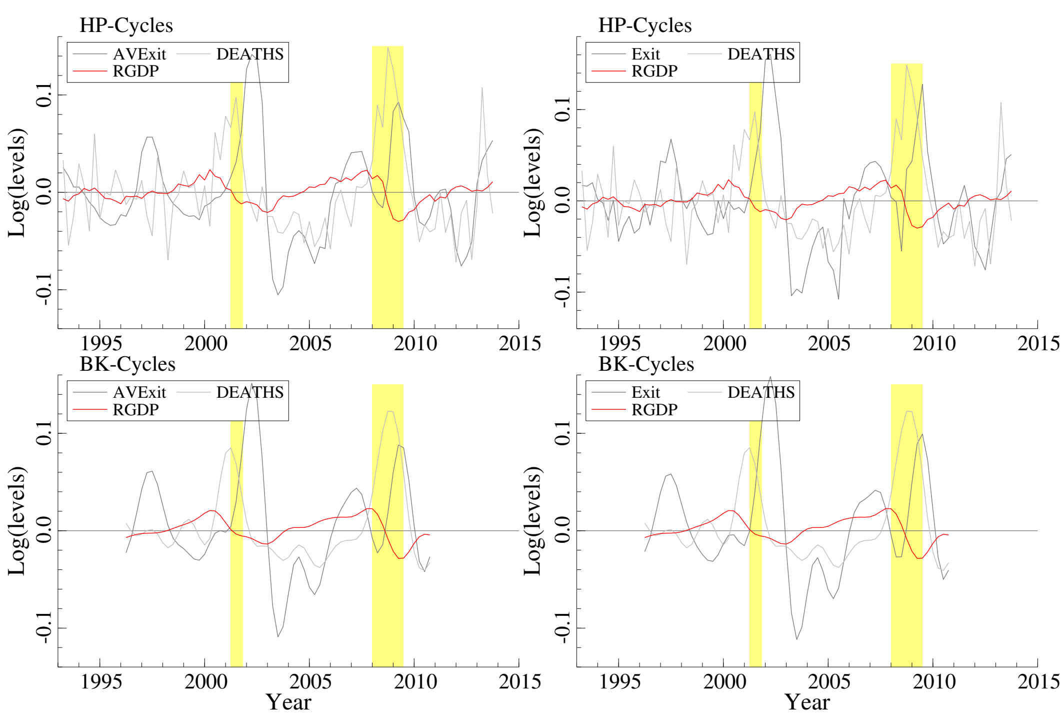
NOTE: This figure shows the business cycle analysis of the BLS existing quarterly series of ENTRY – in panel (a) – and EXIT – in panel (b). Left (right) panels plot the trend (cycle) component extracted from the series in logarithms. Upper (lower) panels refer to HP (BK) filter. The cyclical components extracted by the HP filter assume penalization parameter $\lambda = 1600$ and the ones obtained by BK filter assume parameters $\lambda_0 = 1.5$, $\lambda_1 = 8$ and $K = 12$ (that is, $3 \times s$). Colored bands correspond to the NBER recessions. Software used: OxMetrics.

Figure 7: The cycle component: comparison between different proxies.

(a) ENTRY

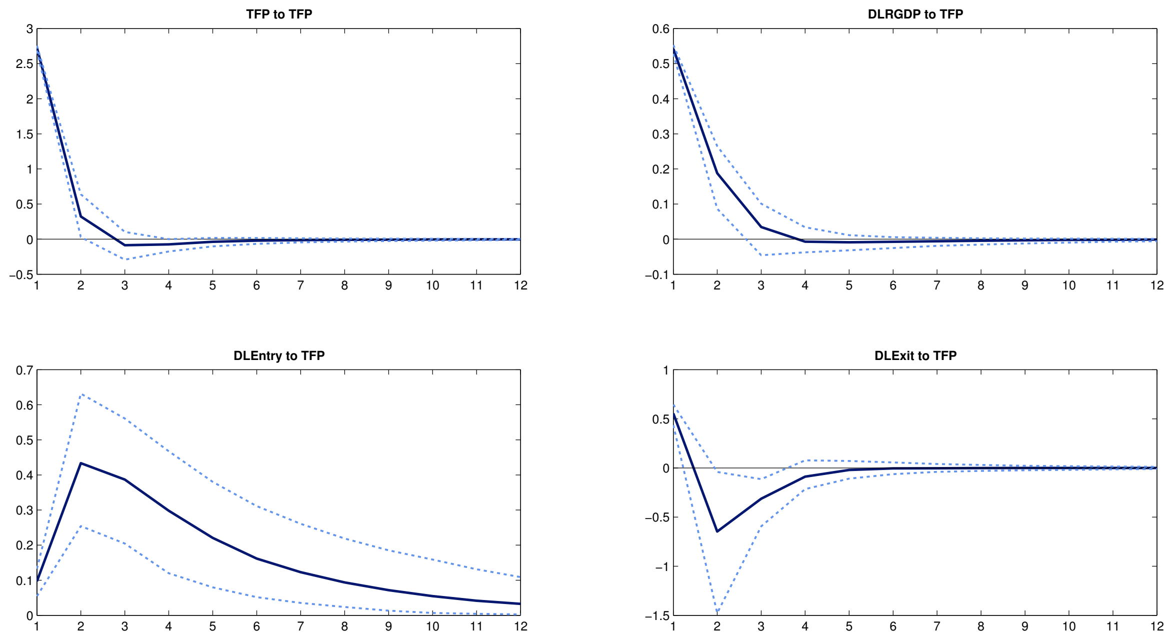


(b) EXIT



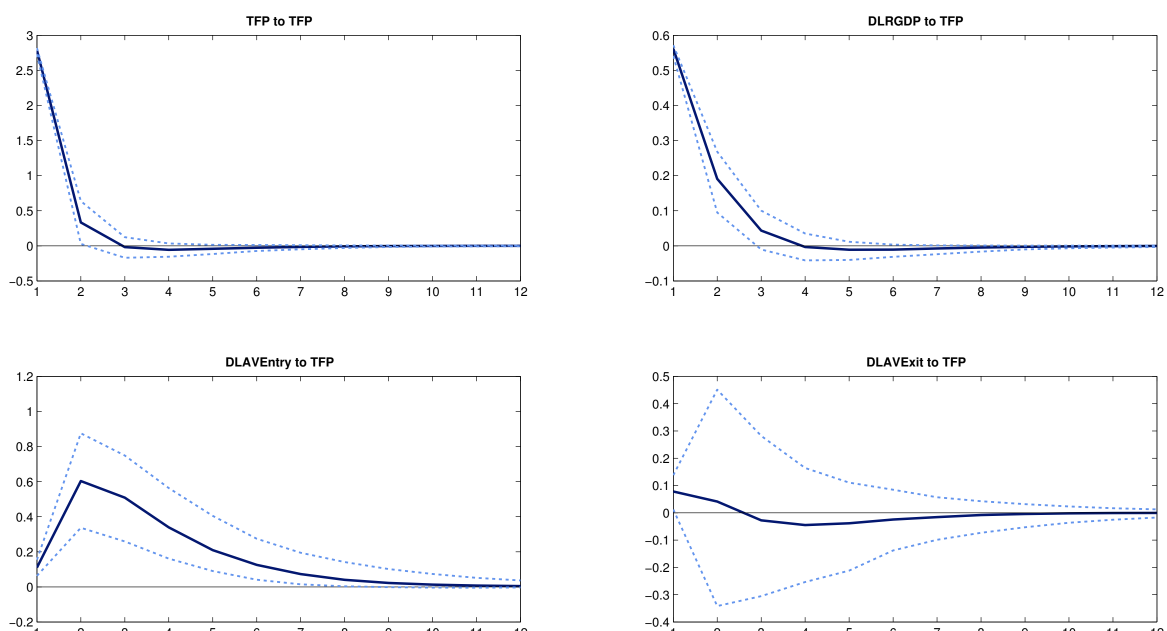
NOTE: This figure shows the cycle of BLS proxies of ENTRY – in panel (a) – and EXIT – in panel (b) for different filters. Namely, the upper (lower) panels refer to the HP (BK) filter. On the other side, left (right) panels compares these proxies with combination (univariate) measures of ENTRY/EXIT. The cyclical components extracted by the HP filter assume penalization parameter $\lambda = 1600$ and the ones obtained by BK filter assume parameters $\lambda_0 = 1.5$, $\lambda_1 = 8$ and $K = 12$ (that is, $3 \times s$). Colored bands correspond to the NBER recessions. Software used: OxMetrics.

Figure 8: Impulse-response functions of M1



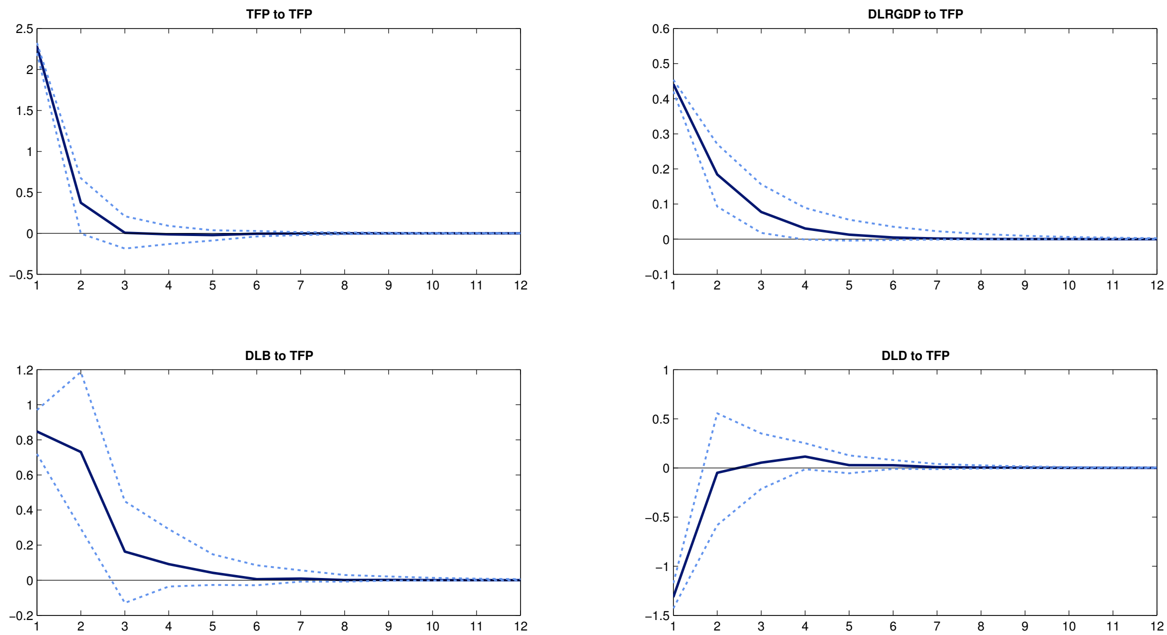
NOTE: This figure plots the values of the IRFs corresponding to equation (19) for M1 – and, specifically, to equation (19) – (Y-axis) for a time horizon of 12 periods (X-axis). Number of draws for bootstrap confidence bands: 300. Software used: MATLAB R2009b.

Figure 9: Impulse-response functions of M2



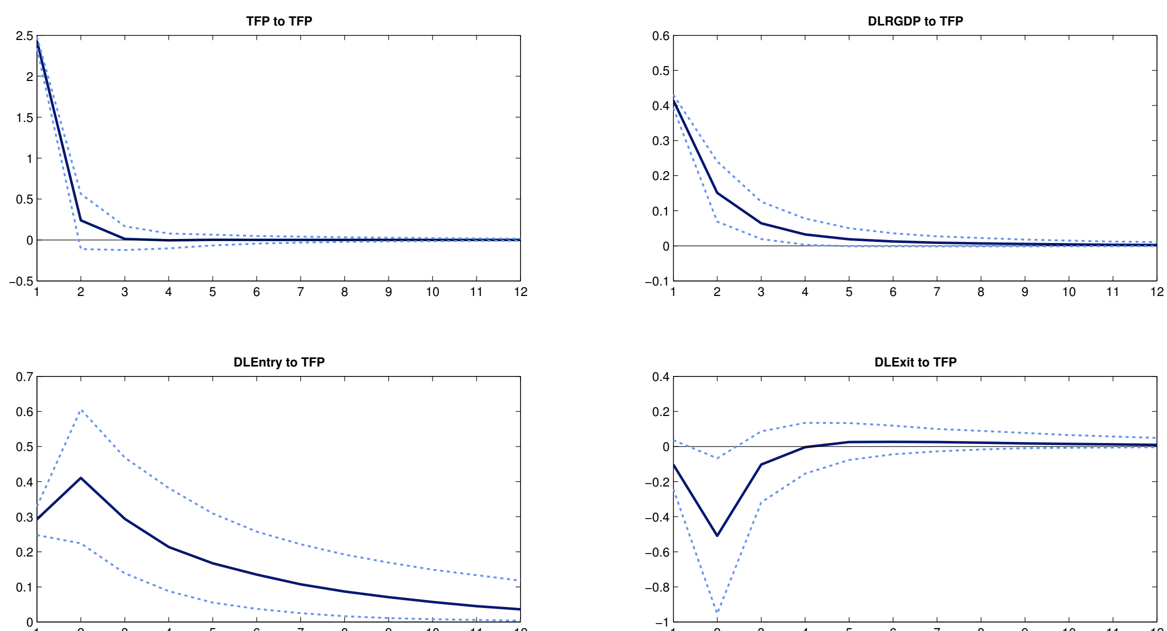
NOTE: This figure plots the values of the IRFs corresponding to equation (19) for M2 (Y-axis) for a time horizon of 12 periods (X-axis). Number of draws for bootstrap confidence bands: 300. Software used: MATLAB R2009b.

Figure 10: Impulse-response functions of M3



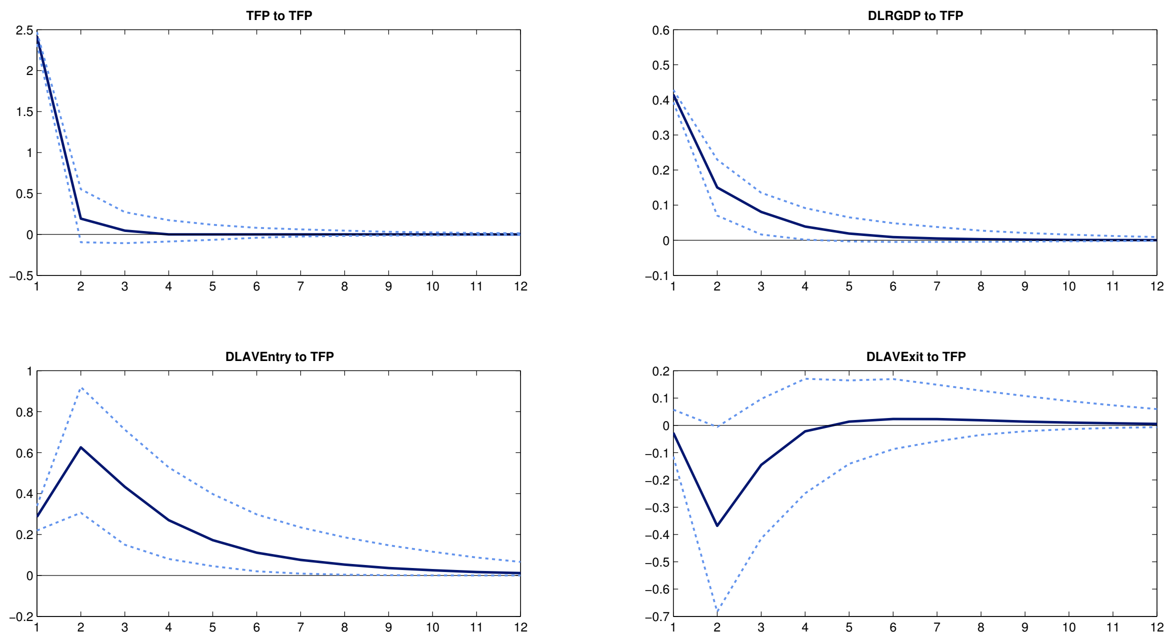
NOTE: This figure plots the values of the IRFs corresponding to equation (19) for M3 (Y-axis) for a time horizon of 12 periods (X-axis). Number of draws for bootstrap confidence bands: 300. Software used: MATLAB R2009b.

Figure 11: Impulse-response functions of M4



NOTE: This figure plots the values of the IRFs corresponding to equation (19) for M4 (Y-axis) for a time horizon of 12 periods (X-axis). Number of draws for bootstrap confidence bands: 300. Software used: MATLAB R2009b.

Figure 12: Impulse-response functions of M5



NOTE: This figure plots the values of the IRFs corresponding to equation (19) for M5 (Y-axis) for a time horizon of 12 periods (X-axis). Number of draws for bootstrap confidence bands: 300. Software used: MATLAB R2009b.

Table 1: Correlations between disaggregated series and 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	

NOTE: This table reports the correlation between the series resulting from disaggregation method accrding to (1) and the method (8) with the specifications (3) and (4) and the existing quarterly data downloaded from BLS.

Table 2: Estimated model parameters of UCM 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	30,716.87	1,599.51	19.15	49,024.53	1,780.91	27.53
		x_2	41.89	18.15	2.31	119.12	20.20	5.90
		x_3	-409,716.78	533,143.27	-0.77	689,839.94	552,907.00	1.25
		x_4	769,250.14	515,659.34	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	32,623.80	1,556.93	20.95	64,365.34	2,077.61	30.98
		x_2	33.73	18.90	1.78	163.93	24.08	6.81
		x_3	-143,156.25	552,615.74	-0.26	36,358.01	479,154.96	0.08
LR-statistic			**4.3471**			1.5835		

NOTE: This table reports the estimates of parameters of equations (3) and (4). Regressors $x_1, \ldots, x_4$ correspond to the deterministic component included in the model, see Remark 2; all the diffuse effects x_t are reported, jointly with their standard deviation and t-statistics. The log-Likelihood corresponds to diffuse profile (14) deriving from estimated parameters in (12). The one-degree-of-freedom Likelihood-ratio test corresponds to the null hypothesis that $\phi = 0$ versus the alternative hypothesis that $\phi \neq 0$ in condition (5); rejection of null hypothesis leads to ADL(1,1) model; the bold is used to indicate that statistic is rejected at 10% of significance.

Table 3: Stylized Business Cycle Facts

Data Types	Series	Lead	Lag	Sign
Disaggregated (Univariate UCM)	Entry-BK		3	+
	Entry–HP		3	+
	Entry–LDiff	8		+
	Exit–BK	6		-
	Exit–HP		8	+
	Exit–LDiff		1	-
Disaggregated (Combined UCM)	AVEntry-BK		3	+
	AVEntry–HP		3	+
	AVEntry–LDiff		1	+
	AVExit–BK	6		-
	AVExit–HP	7		-
	AVExit–LDiff	7		-
BLS	Openings-BK		0	+
	Openings–HP		0	+
	Openings–LDiff		0	+
	Closings–BK		5	+
	Closings–HP		5	+
	Closings–LDiff		4	+
BLS	Births-BK		0	+
	Births–HP		0	+
	Births–LDiff		1	+
	Deaths–BK	4		-
	Deaths–HP	5		+
	Deaths–LDiff		2	+

NOTE: This table reports the cross-correlation analysis of the cyclical component of different measure of ENTRY and EXIT respect to the RGDP. "Lead" ("Lag") indicates the point where the cross-correlation function – in absolute value – of ENTRY (EXIT) and RGDP reach its maximum; a zero value of lag/lead indicate that the series is a coincident indicator of the business cycle. The sign "+" ("-") means that the maximum cross-correlation in the indicated lead/lag is positive (negative). The cyclical components extracted by the HP filter assume penalization parameter $\lambda = 1600$ and the ones obtained by BK filter assume parameters $\lambda_0 = 1.5$, $\lambda_1 = 8$ and $K = 12$ (that is, $3 \times s$).

Table 4: VAR Models description

Model	Sample	Y_t (in Δlog)
M1	1977:Q1–2013:Q4	[RGDP Entry Exit TFP]'
M1b		[RGDP INFL Entry Exit TFP]'
M2		[RGDP AVEntry AVExit TFP]'
M2b		[RGDP INFL AVEntry AVExit TFP]'
M4	1992:Q3–2013:Q4	[RGDP Entry Exit TFP]'
M4b		[RGDP INFL Entry Exit TFP]'
M5		[RGDP AVEntry AVExit TFP]'
M5b		[RGDP INFL AVEntry AVExit TFP]'
M3	1993:Q3–2013:Q4	[RGDP BIRTHS DEATHS TFP]'
M3b		[RGDP INFL BIRTHS DEATHS TFP]'
R1		[RGDP BIRTHS DEATHS TFPu]'
R1b		[RGDP INFL BIRTHS DEATHS TFPu]'
R2		[RGDP BIRTHS DEATHS LABPROD]'
R2b		[RGDP INFL BIRTHS DEATHS LABPROD]'

NOTE: This table reports a list of definitions of vector $\mathbf{y_t}$ to be modelled by the VAR system (16) for structural analysis. Namely, M1–M5 are shown in the course of the paper, while M1b–M5b and R1(b)–R2(b) are shown in Supplement; the label "R" adopted for R1 and R2 indicates that the VAR model corresponding to these definition of the equation system coincides with the theoretical one in Rossi (2015). All the series are expressed in growth rates; "INFL" (inflation) is computed as $\Delta \log CPI_t$

Table 5: VAR estimates of M1

Parameter	Summary Statistics
T	146
p	20
Log-Lik	-1,575.5608
R^2(LR)	0.7875

Estimation

Variable	ΔLRGDP			ΔLEntry			ΔLExit			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.0041	0.0009	0.0000	-0.0041	0.0019	0.0402	0.0105	0.0074	0.1632	0.0072	0.0034	0.0357
ΔLRGDP(1)	0.3884	0.1219	0.0018	0.5987	0.2900	0.0408	-1.2767	0.8379	0.1298	-0.0261	0.4431	0.9531
ΔLEntry(1)	-0.0189	0.0300	0.5282	0.8166	0.0665	0.0000	-0.0652	0.1558	0.6760	-0.0672	0.1018	0.5098
ΔLExit(1)	0.0374	0.0146	0.0117	0.1027	0.0316	0.0014	0.1176	0.1010	0.2461	0.1460	0.0567	0.0112
ΔLTFP(1)	-0.0126	0.0268	0.6374	-0.0077	0.0019	0.9219	0.0250	0.1700	0.8833	0.1220	0.1263	0.3355
σ		0.0071			0.0159			0.0409			0.0277	
RSS		0.0071			0.0358			0.2364			0.1083	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degree of freedom)	Statistic (value)	*p*-value
Godfrey	Autocorrelation	F(32,481)	4.9947	0.0000**
Doornik-Hansen	Normality	$\chi^2(8)$	108.3200	0.0000**
White	Heteroskedasticity	F(32,495)	2.7309	0.0000**
RESET	Correct Specification	F(32,481)	1.6192	0.0190*

NOTE: This table reports the estimates of M1according to the unrestriced VAR(1) model 16. In the upper part, we provide the number of observations (T), the number of parameters of the system (p), the Log-Likelihood ($\Lambda = -T/2log|\hat{\mathbf{\Omega}}| - Tp/2(1 + log2\pi)$) and the R^2 based on the Likelihood-principle ($R^2(LR) = 1 - |\hat{\mathbf{\Omega}}|\hat{\mathbf{\Omega}}_\mathbf{0}$, with $\hat{\mathbf{\Omega}}_\mathbf{0}$ denoting the residual variance under the null hypothesis of no unrestricted elements in the system). The central part shows the estimates of $\boldsymbol{A_1}$ parameter matrix, jointly with its heteroskedastic-consistent standard errors and t-values, the standard errors ($\tilde{\sigma}$) and residual sum of squares (RSS) for each equation of the system. The lower part shows four standard diagnostic tests on the whole vector system, namely: (i) a Hendry and Doornik-type test for the hypothesis of no autocorrelation on residuals of the model; (ii) the Doornik and Hansen (2008) test for the hypothesis that the skewness and kurtosis of estimated residuals correspond to those of a multivariate normal distribution; (iii) the White (1980) test for the null hypothesis of unconditional homoscedasticity against the alternative that the variance of the vector error process depends on the regressors and their square; and (iv) the Ramsay (1969) Regression Specification Test (RESET) for the null hypothesis of correct specification of the original model against the alternative that powers of $\hat{\boldsymbol{y}}_{\boldsymbol{t}}$ have been omitted (here we assume only squares and cubes). Rejection of one or more of these hypotheses indicates a potential pitfall of the model in describing efficiently the data. "**" indicates rejection at 1%, "*" rejection at 5%. Software used for calculation: PC-Give; for further details, see the PC-Give Help and Hendry and Doornik (2009).

Table 6: VAR estimates of M2

Parameter	Summary Statistics
T	146
p	20
Log-Lik	1611.9161
R^2(LR)	0.7758

Estimation

Variable	ΔLRGDP			ΔLAVEntry			ΔLAVExit			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.0041	0.0009	0.0000	-0.0036	0.0029	0.2138	0.0024	0.0037	0.5175	0.0075	0.0034	0.0294
ΔLRGDP(1)	0.3812	0.1185	0.0016	0.4004	0.3480	0.2467	-0.3192	0.4130	0.4407	-0.0698	0.4376	0.8734
ΔLAVEntry(1)	-0.0399	0.0280	0.1566	0.6949	0.0798	0.0000	-0.0329	0.0887	0.7106	-0.1193	0.1026	0.2466
ΔLAVExit(1)	-0.0080	0.0228	0.7264	0.0443	0.0434	0.3092	0.5868	0.0800	0.0000	0.0111	0.0896	0.9013
ΔLTFP(1)	0.03520	0.0265	0.8946	-0.1099	0.0727	0.1332	-0.0719	0.1206	0.5520	0.1589	0.01156	0.1711
σ		0.0072			0.0196			0.0245			0.0283	
RSS		0.0074			0.0537			0.0851			0.1133	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degree of freedom)	Statistic (value)	p-value
Godfrey	Autocorrelation	F(80,467)	2.2731	0.0000**
Doornik-Hansen	Normality	$\chi^2(8)$	99.1500	0.0000**
White	Heteroskedasticity	F(32,495)	1.6770	0.0127*
RESET	Correct Specification	F(32,481)	2.0567	0.0000**

NOTE: This table reports the estimates of M2 according to the unrestriced VAR(1) model 16. In the upper part, we provide the number of observations (T), the number of parameters of the system (p), the Log-Likelihood ($\Lambda = -T/2log|\hat{\mathbf{\Omega}}| - Tp/2(1 + log2\pi)$) and the R^2 based on the Likelihood-principle ($R^2(LR) = 1 - |\hat{\mathbf{\Omega}}|\hat{\mathbf{\Omega}}_\mathbf{0}$, with $\hat{\mathbf{\Omega}}_\mathbf{0}$ denoting the residual variance under the null hypothesis of no unrestricted elements in the system). The central part shows the estimates of $\boldsymbol{A_1}$ parameter matrix, jointly with its heteroskedastic-consistent standard errors and t-values, the standard errors ($\tilde{\sigma}$) and residual sum of squares (RSS) for each equation of the system. The lower part shows four standard diagnostic tests on the whole vector system, namely: (i) a Hendry and Doornik-type test for the hypothesis of no autocorrelation on residuals of the model; (ii) the Doornik and Hansen (2008) test for the hypothesis that the skewness and kurtosis of estimated residuals correspond to those of a multivariate normal distribution; (iii) the White (1980) test for the null hypothesis of unconditional homoscedasticity against the alternative that the variance of the vector error process depends on the regressors and their square; and (iv) the Ramsay (1969) Regression Specification Test (RESET) for the null hypothesis of correct specification of the original model against the alternative that powers of $\hat{\boldsymbol{y}}_{\boldsymbol{t}}$ have been omitted (here we assume only squares and cubes). Rejection of one or more of these hypotheses indicates a potential pitfall of the model in describing efficiently the data. "**" indicates rejection at 1%, "*" rejection at 5%. Software used for calculation: PC-Give; for further details, see the PC-Give Help and Hendry and Doornik (2009).

Table 7: VAR estimates of M3

Parameter	Summary Statistics
T	81
p	20
Log-Lik	848.7591
R^2(LR)	0.6877

Variable	ΔLRGDP			ΔLB			ΔLD			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.0034	0.0011	0.0047	-0.0110	0.0052	0.0384	-0.0044	0.0061	0.4746	0.0103	0.0044	0.0239
ΔLRGDP(1)	0.4699	0.1553	0.0034	1.7355	0.8162	0.0367	2.4548	0.8586	0.0055	0.0567	0.6472	0.9304
ΔLB(1)	0.0022	0.0147	0.8801	-0.3143	0.1214	0.0115	-0.2355	0.1280	0.0696	-0.0573	0.0621	0.3589
ΔLD(1)	-0.0015	0.0139	0.9137	0.1623	0.1021	0.1160	-0.3294	0.1190	0.0071	-0.1447	0.0613	0.0208
ΔLTFP(1)	-0.0116	0.0318	0.7162	0.2029	0.2348	0.3902	-0.6141	0.2293	0.0091	0.0875	0.1587	0.5827
σ		0.0058			0.0363			0.0387			0.0234	
RSS		0.0026			0.1002			0.1143			0.0419	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degrees of freedom)	Statistic (value)	p-value
Godfrey	Autocorrelation	F(80,211)	1.6587	0.0022**
Doornik-Hansen	Normality	$\chi^2(8)$	10.940	0.2051
White	Heteroskedasticity	F(32,256)	0.9623	0.5298
RESET	Correct Specification	F(32,241)	1.2934	0.1436

NOTE: This table reports the estimates of M3 according to the unrestriced VAR(1) model 16. In the upper part, we provide the number of observations (T), the number of parameters of the system (p), the Log-Likelihood ($\Lambda = -T/2log|\hat{\mathbf{\Omega}}| - Tp/2(1 + log2\pi)$) and the R^2 based on the Likelihood-principle ($R^2(LR) = 1 - |\hat{\mathbf{\Omega}}|\hat{\mathbf{\Omega}}_\mathbf{0}$, with $\hat{\mathbf{\Omega}}_\mathbf{0}$ denoting the residual variance under the null hypothesis of no unrestricted elements in the system). The central part shows the estimates of $\boldsymbol{A_1}$ parameter matrix, jointly with its heteroskedastic-consistent standard errors and t-values, the standard errors ($\tilde{\sigma}$) and residual sum of squares (RSS) for each equation of the system. The lower part shows four standard diagnostic tests on the whole vector system, namely: (i) a Hendry and Doornik-type test for the hypothesis of no autocorrelation on residuals of the model; (ii) the Doornik and Hansen (2008) test for the hypothesis that the skewness and kurtosis of estimated residuals correspond to those of a multivariate normal distribution; (iii) the White (1980) test for the null hypothesis of unconditional homoscedasticity against the alternative that the variance of the vector error process depends on the regressors and their square; and (iv) the Ramsay (1969) Regression Specification Test (RESET) for the null hypothesis of correct specification of the original model against the alternative that powers of $\hat{\boldsymbol{y}}_{\boldsymbol{t}}$ have been omitted (here we assume only squares and cubes). Rejection of one or more of these hypotheses indicates a potential pitfall of the model in describing efficiently the data. "**" indicates rejection at 1%, "*" rejection at 5%. Software used for calculation: PC-Give; for further details, see the PC-Give Help and Hendry and Doornik (2009).

Table 8: VAR estimates of M4

Parameter	Summary Statistics
T	85
p	20
Log-Lik	990.3584
R^2(LR)	0.8400

Variable	ΔLRGDP			ΔLEntry			ΔLExit			ΔLTFP		
	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob	Coeff.	HCSE	t-prob
CONST	0.0035	0.1534	0.0025	-0.0010	0.2899	0.5593	0.0082	0.0064	0.2005	0.0092	0.0046	0.0468
ΔLRGDP(1)	0.4346	0.0297	0.0058	0.0908	0.0673	0.7548	-1.0902	0.8062	0.1801	-0.1486	0.7075	0.8341
ΔLEntry(1)	0.0283	0.0223	0.3438	0.8096	0.0379	0.0000	0.3215	0.1929	0.0995	0.0304	0.1509	0.8407
ΔLExit(1)	0.0041	0.0296	0.8537	0.0965	0.0824	0.0129	0.1628	0.1381	0.2419	0.0138	0.0911	0.8792
ΔLTFP(1)	-0.0098	0.0011	0.7392	0.0703	0.0018	0.3963	-0.0535	0.1982	0.7875	0.1535	0.1603	0.3411
σ		0.0057			0.0111			0.0333			0.0249	
RSS		0.0026			0.0098			0.0888			0.0496	

Diagnostic Tests on the Vector System

Test type	Test Hypothesis	Statistic used (degree of freedom)	Statistic (value)	*p*-value
Godfrey	Autocorrelation	F(80,227)	2.4472	0.0000**
Doornik-Hansen	Normality	$\chi^2(8)$	17.642	0.0241*
White	Heteroskedasticity	F(32,270)	1.2115	0.2085
RESET	Correct Specification	F(32,256)	1.4069	0.0792

NOTE: This table reports the estimates of M4 according to the unrestriced VAR(1) model 16. In the upper part, we provide the number of observations (T), the number of parameters of the system (p), the Log-Likelihood ($\Lambda = -T/2log|\hat{\mathbf{\Omega}}| - Tp/2(1 + log2\pi)$) and the R^2 based on the Likelihood-principle ($R^2(LR) = 1 - |\hat{\mathbf{\Omega}}|\hat{\mathbf{\Omega_0}}$, with $\hat{\mathbf{\Omega_0}}$ denoting the residual variance under the null hypothesis of no unrestricted elements in the system). The central part shows the estimates of $\boldsymbol{A_1}$ parameter matrix, jointly with its heteroskedastic-consistent standard errors and t-values, the standard errors ($\tilde{\sigma}$) and residual sum of squares (RSS) for each equation of the system. The lower part shows four standard diagnostic tests on the whole vector system, namely: (i) a Hendry and Doornik-type test for the hypothesis of no autocorrelation on residuals of the model; (ii) the Doornik and Hansen (2008) test for the hypothesis that the skewness and kurtosis of estimated residuals correspond to those of a multivariate normal distribution; (iii) the White (1980) test for the null hypothesis of unconditional homoscedasticity against the alternative that the variance of the vector error process depends on the regressors and their square; and (iv) the Ramsay (1969) Regression Specification Test (RESET) for the null hypothesis of correct specification of the original model against the alternative that powers of $\hat{\boldsymbol{y}}_{\boldsymbol{t}}$ have been omitted (here we assume only squares and cubes). Rejection of one or more of these hypotheses indicates a potential pitfall of the model in describing efficiently the data. "**" indicates rejection at 1%, "*" rejection at 5%. Software used for calculation: PC-Give; for further details, see the PC-Give Help and Hendry and Doornik (2009).

Technical Supplement
to
"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*

FIRST VERSION: May 2016
THIS VERSION: June 2017

1 Introduction

This Supplement aims to provide more details in the statistical tools used in Rossi and Zanetti Chini (2017) (RZC, henceforth) and, secondly, to corroborate the empirical evidence. The next Section 2 discusses briefly the estimation method of the Proietti (2006) disaggregation models used in our paper and the Markov-Switching VAR (MS-VAR) model used to investigate the possible presence nonlinearity in our data. Finally, Section 3 describes further diagnostic and robustness checks.

2 Statistical tools

2.1 The Augmented Kalman Filter and Smoother

Under the initial conditions mentioned in page 10 of RZC, 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),

$$v_t = y_t^c - z^{*'} a_t^{*'}, \qquad V_t' = z^{*'} A_t^*, \tag{1}$$

$$f_t = z^{*'} P_t^* z^*, \qquad K_t = T_t^* P_t^* z^* / f_t, \tag{2}$$

$$\alpha_{t+1}^* = T_{t+1}^* a_t^* + K_t v_t, \qquad A_{t+1}^* = W_{t+1}^* A_t^* T_{t+1}^* + K_t V_t' \tag{3}$$

$$P_{t+1}^* = T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'} - K_t K_t' f_t \qquad s_{t+1} = s_t V_t v_t / f_t \tag{4}$$

$$q_{t+1} = q_t + v_t^2 / f_t \qquad d_{t+1} = d_t + \ln f_t \tag{5}$$

$$S_{t+1} = S_t + V_t V_t' / f_t \qquad t = 1, \ldots, T \tag{6}$$

$$\tag{7}$$

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

$$\alpha_{t+1}^* = T_{t+1}^* a_t^* \qquad A_{t+1}^* = W_{t+1}^* A_t^* T_{t+1}^* \tag{8}$$

$$P_{t+1}^* = T_{t+1}^* P_t^* T_{t+1}^{*'} + H_t^* H_t^{*'} \qquad s_{t+1} = s_t \tag{9}$$

$$q_{t+1} = q_t, \qquad S_{t+1} = S_t \tag{10}$$

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:

$$\hat{\alpha}_t | t = a_t^* - A_t^* S_t^{-1} s_t + P_t^* z^* \hat{v}_t / f_t, \qquad P_{t|t}^* = P_t^* + A_t^* S_t^{-1} A_t^{*'} - P_t^* z^* z' P_t^* / f_t \tag{11}$$

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),

$$r_{t-1} = z^* v_t / f_t + (T_{t+1} + K_t z^{*'}) r_t, \qquad R_{t-1} = z^* V_t^{'} / f_t + (T_{t+1} - K_t z^{*'}) R_t \quad (12)$$

$$N_{t-1} = z^* z^{*'} / f_t + (T_{t+1} - K_t z^{*'}) N_t (T_{t+1} - K_t z^{*'})', \quad (13)$$

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

$$r_{t-1} = T_{t+1} r_t, \qquad R_{t-1} = T_{t+1} R_t, \qquad R_{t-1} = T_{t+1} N_t T_{t+1}^{'}. \quad (14)$$

The smoothed estimates are obtained as

$$\hat{\alpha}_{t|t}^* = a_t^* + A_t^* \hat{\beta} + P_t^* (r_{t-1} + R_{t-1} \hat{\beta}) \qquad P_{t|t}^* = P_t^* + A_t^* S_{n+1}^{-1} A_t^{*'} - P_t^* N_{t-1} P_t^* \quad (15)$$

2.2 MS-VAR

Markov-switching vector autoregressions can be considered as generalizations of the basic finite order VAR model of order p. The general idea behind this class of models is that the parameters of the underlying data generating process of the observed time series vector y_t depend upon the unobservable regime variable s_t, which represents the probability of being in a different state of the world. Thus, the p-th order autoregression for the K-dimensional time series vector $y_t = (y_{1t}, \ldots, y_{Kt})$, $t = 1, \ldots, T$,

$$\boldsymbol{y}_t = \boldsymbol{\mu}(s_t) + \boldsymbol{A}_1(s_t) y_{t-1} + \cdots + \boldsymbol{A}_p(s_t) y_{t-p} + \boldsymbol{u}_t, \quad (16)$$

where A_j is parameter matrix $s_t \in (1, \ldots, M)$ is governed by a discrete time, a discrete state, and a irreducible ergodic M state Markov process with the transition probabilities matrix defines as:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix} \tag{17}$$

with p_{ij} the probability of switching from state i to state j, that is

$$p_{ij} = Pr(s_{t+1} = j | s_t = i), \quad \sum_{j=1}^{M} p_{ij} = 1 \ \forall i, j \in \{1, \ldots, M\} \tag{18}$$

Denoting $A(L) = \mathbf{I}_K - \mathbf{\Pi}_1 L - \ldots, -\mathbf{\Pi}_p L^p$ as the $(K \times K)$ dimensional lag polynomial, we assume that there are no roots on or inside the unit circle $|\mathbf{\Pi}(z)| \neq 0$ for $|z| \leq 1$ where L is the lag operator, so that $y_{t-j} = L^j y_t$. If a normal distribution of the error is assumed, $u_t \sim NID(0, \Sigma(s_t))$, equation (16) is known as the intercept form of a stable Markov Switching Gaussian VAR(p) model.

This model (16) is the most general case of MS-AR family. In our application, for example, we avoid to assume that also the intercept is regime-switching, that is we impose $\boldsymbol{\mu}(s_t) = \mu \, \forall \, s \in 1, \ldots, M$. For further details on the MS-VAR family, see Krolzig (2013).

3 Further Results

3.1 Residual Analysis

Let start with the residual diagnostics of the VAR models used in RZC. The statistical properties of the M1–M5 can be appreciated in Figures 1 – 5, 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, in order to have a more precise image of each quantile, the QQ plot (fourth column). It is easy to notice the high kurtosis of the ENTRY and EXIT variables (univariate and combined) with respect to RGDP. Anyway, the comparison of quantiles with of the estimated model are considerably similar to the standard Normal ones.

3.2 Additional Checks

The four-variable VAR models above presented, despite their simplistic formulation, contains all the elements to estimate the effects of a structural shock on the firms dynamics over the business cycle. Anyway, the economic theory is quite less parsimonious in its formulation. In particular, according to Rossi (2015), the inflation plays a central role in the propagation of the shock from the firms to the overall business cycle.

Moreover, as mentioned in RZC, a second potential source of over-simplification in the analysis is in the linearity assumption of the VAR structure. In practice, it is very-well known in empirical literature that business cycle is asymmetric, see Burns and Mitchell (1946); Neftçi (1984); Sichel (1993) *inter alia*. This is particularly true if considering that the UCM disaggregated series are based on the Industrial Production, which is one of the most classical example of nonlinear time series in Economics, see Anderson and Teräsvirta (1992) *inter alia* and the results in Zanetti Chini (2017) in a univariate scenario. This means that a different time series model might be assumed to model the data. Thus, we perform several additional checks to overcome to these idiosyncrasies.

To this aim, we repeat our estimation exercise by assuming that RGDP, and ENTRY and EXIT in their different proxies, follow a two-state Markov-Switching Autoregressive (MS-AR) model, whereas the State 1 is the Expansion and State 2 is Recession. Table 1 reports the results for four bivariate models on RGDP and, respectively, EXIT (M1a), ENTRY (M2a) CLOSINGS (M3a) and OPENINGS (M4a).

The asymmetry in the cycle is represented by the difference in the expected duration of the two regimes: only one year and a quarter (a half) of Recession in average for M1a (M3a) that became almost two (two and half) in M2a (M4a). Alternatively, the same feature can be measured by looking at the transition probability matrix: only 0.14 of probability to passing from Expansion to Recession in M1a, and 0.67 in M2; if we change the proxies - that is, the case of M3a and M4a - the conditional State probabilities to pass from State 1 to State 2 (0.77 and 0.41, respectively). Finally, it is interesting to notice that only for M1a the correlation between RGDP and EXIT changes in sign when moving from positive in state of Expansion (0.05) to negative the state of Recession (-0.18), while in all other cases the only difference is in the magnitude. This seems to suggest that, in terms of stylized facts, our disaggregated data are more coherent with the theory.

Nevertheless, a deeper investigation on the graphical output of the MS-AR estimation plotted in Figures 6 reveals that, with the relevant exception of two short periods - the Grate Recession and first 80s - the disaggregated series are quite stable. The only model witch present a more pronounced nonlinear path is M3a, as expected in light of the previous analysis. This leads us to abandon the inspection of the non-linear scenario to explore the role of different variables and shocks in the traditional linear SVAR.

To this aim, we use the model by Rossi (2015) to investigate the effect of two different types of shocks: on utlity-adjusted TFP (TFPu) and on Labor Productivity (LABPROD); these two variables are inserted in four (five) variable VAR with RGDP, BIRTHS and DEATHS (and INFLATION) and the resulting models are labeled R1 (R1b) and R2 (R2b), see also Table 4 in RZC. The IRFs shown in Figures 11 – 14, are consistent with the theoretical ones.

Finally, we run a battery of five-variable VARs similar to M1 – M5 but with the new variable INFL. These are labeled as M1b, M2b, M3b, M4b and M5b to facilitate their comparison with the four-variable models. The corresponding IRF are shown in Figures 7 – 28. We can summarize the results as follows:

(i) The SVAR with short run restrictions are globally consistent and better, in terms of quality of IRF and error bands, than the ones with long-run restrictions;

(ii) the responses of models with univariate disaggregation of ENTRY and EXIT has error bands considerably more tight;

(iii) there is no clear superiority of data from 1993 with respect to our disaggregation from 1977.

References

Anderson, H. and Teräsvirta, T. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, **7** (S1), S119–S136.

Burns, A. and Mitchell, W. (1946). Measuring Business Cycles. National Bureau of Economic Research.

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

Krolzig, H. (2013). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Berlin.

Neftçi, S. (1984). Are Economic Time Series Asymmetric over the Business Cycle? *Journal of Political Economy*, **92**, 307–328.

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 no. 99, University of Pavia.

— and Zanetti Chini, E. (2017). Firms' Dynamics and Business Cycle: New Disaggregated Data. DEM Working Paper No. ???, University of Pavia.

Sichel, D. (1993). Business cycle asymmetry: a deeper look. *Economic Inquiry*, **31** (2), 224–236.

Zanetti Chini, E. (2017). Generalizing Smooth Transition Autoregression. DEM Working Paper No. 138, University of Pavia.

Table 1: MS-VAR Estimates

Feature	M1a	M2a	M3a	M4a
Final Log-likelihood:	−452.72	−341.48	−268.42	−246.48
No. of estimated parameters:	20	20	20	20
No. of Effective Observations:	132	132	82	82
Expected Duration for Regime 1	86.50	122.75	10.09	30.42
Expected Duration for Regime 2	5.03	7.10	1.49	2.41

	M1a				M2a				M3a				M4a			
Parameter	State 1		State 2		State 1		State 2		State 1		State 2		State 1		State 2	
	Value	SE	Value	SE	Value	SE	Value	SE	Value	SE	Value	SE	Value	SE	Value	SE
k_1	0.40	0.07	-0.02	0.31	0.43	0.08	-0.40	0.29	0.52	0.08	0.36	0.74	0.45	<0.01	-0.72	<0.01
k_2	1.13	0.46	-2.84	2.03	-0.24	0.20	0.24	0.90	-1.03	0.74	1.12	1.61	-1.70	<0.01	-0.96	<0.01
ϕ_{11}	0.47	0.07	-0.47	0.28	0.45	0.09	-0.06	0.35	0.27	0.10	0.58	0.42	0.37	<0.01	-0.44	<0.01
ϕ_{12}	0.01	0.01	0.19	0.09	-0.03	0.02	0.04	0.10	0.00	0.01	0.00	0.06	0.01	<0.01	0.00	<0.01
ϕ_{21}	-1.11	0.47	1.82	1.66	-0.52	0.20	0.70	0.05	2.50	0.88	-1.09	0.92	2.79	<0.01	-3.34	<0.01
ϕ_{22}	0.14	0.08	-0.24	0.46	0.19	0.91	1.02	0.24	-0.20	0.12	-0.84	0.13	-0.40	<0.01	0.18	<0.01

Transition Matrix (p-value in brakets)

Final State Probability	$P_{i\|1}$	$P_{i\|2}$	$P_{i\|1}$	$P_{i\|2}$	$P_{i\|1}$	$P_{i\|2}$	$P_{i\|1}$	$P_{i\|2}$
$P_{1\|j}$	0.99 [<0.01]	0.14 [<0.01]	0.90 [<0.01]	0.67 [0.30]	0.90 [<0.01]	0.77 [<0.01]	0.97 [<0.01]	0.41 [<0.01]
$P_{2\|j}$	0.01 [<0.01]	0.86 [<0.01]	0.10 [<0.01]	0.33 [0.30]	0.10 [<0.01]	0.23 [<0.01]	0.03 [<0.01]	0.59 [<0.01]

Correlation Matrix (p-value in brakets)

State 1	1.00 [<0.01]	0.05 [<0.01]	1.00 [<0.01]	0.19 [<0.01]	1.00 [<0.01]	−0.12 [<0.01]	1.00 [<0.01]	0.19 [<0.01]
	0.05 [<0.01]	1.00 [<0.01]	0.19 [<0.01]	1.00 [<0.01]	−0.12 [<0.01]	1.00 [<0.01]	0.19 [<0.01]	1.00 [<0.01]
State 2	1.00 [<0.01]	−0.18 [<0.01]	1.00 [<0.01]	0.04 [<0.01]	1.00 [<0.01]	−0.99 [<0.01]	1.00 [<0.01]	0.99 [<0.01]
	−0.18 [<0.01]	1.00 [<0.01]	0.04 [<0.01]	1.00 [<0.01]	−0.99 [<0.01]	1.00 [<0.01]	0.99 [<0.01]	1.00 [<0.01]

Figure 1: The estimated VAR model residuals for M1

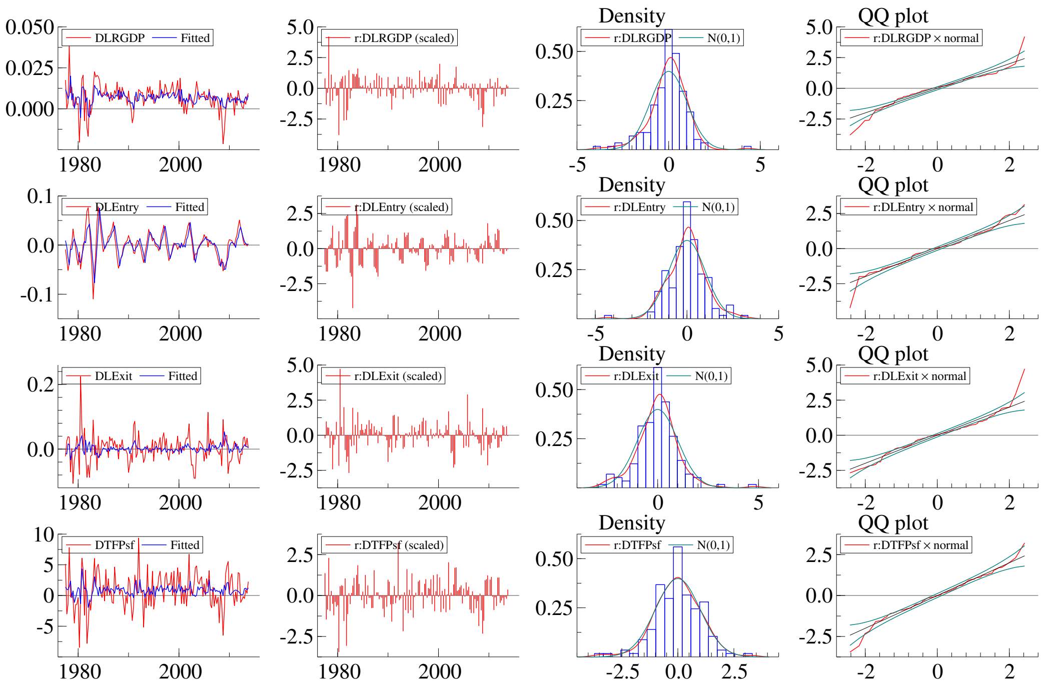


Figure 2: The estimated VAR model residuals for M2

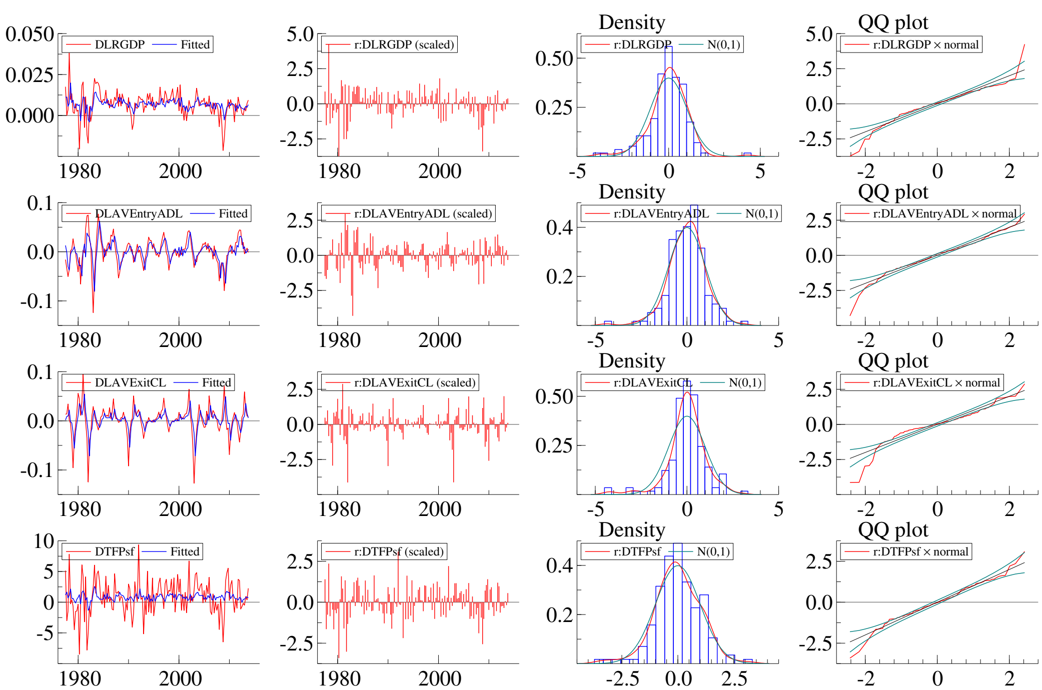


Figure 3: The estimated VAR model residuals for M3

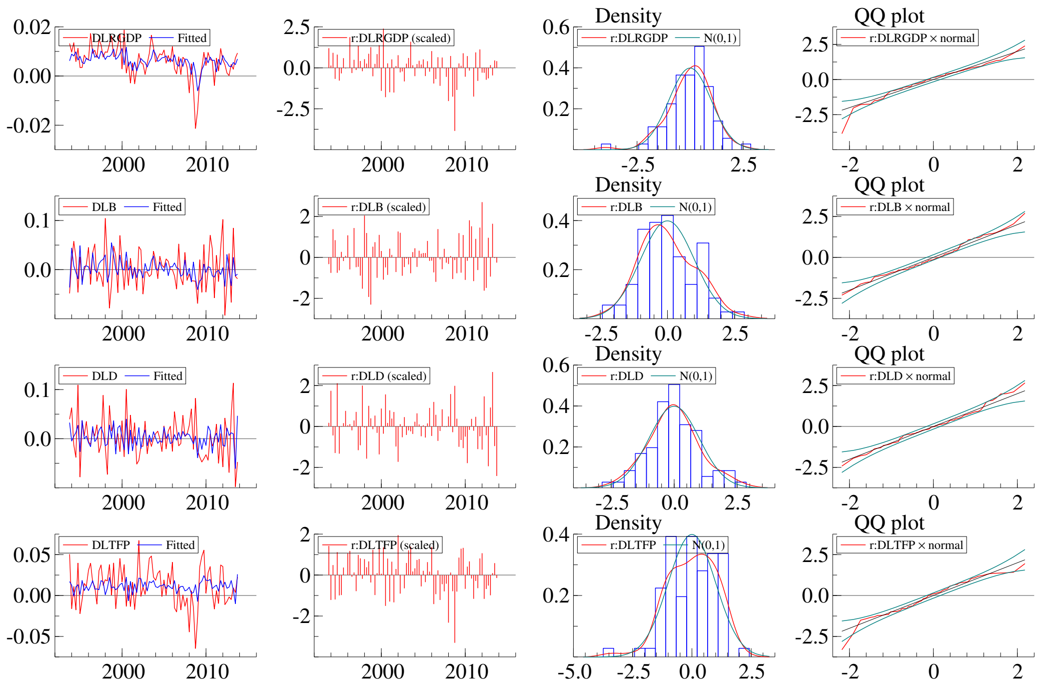


Figure 4: The estimated VAR model residuals for M4

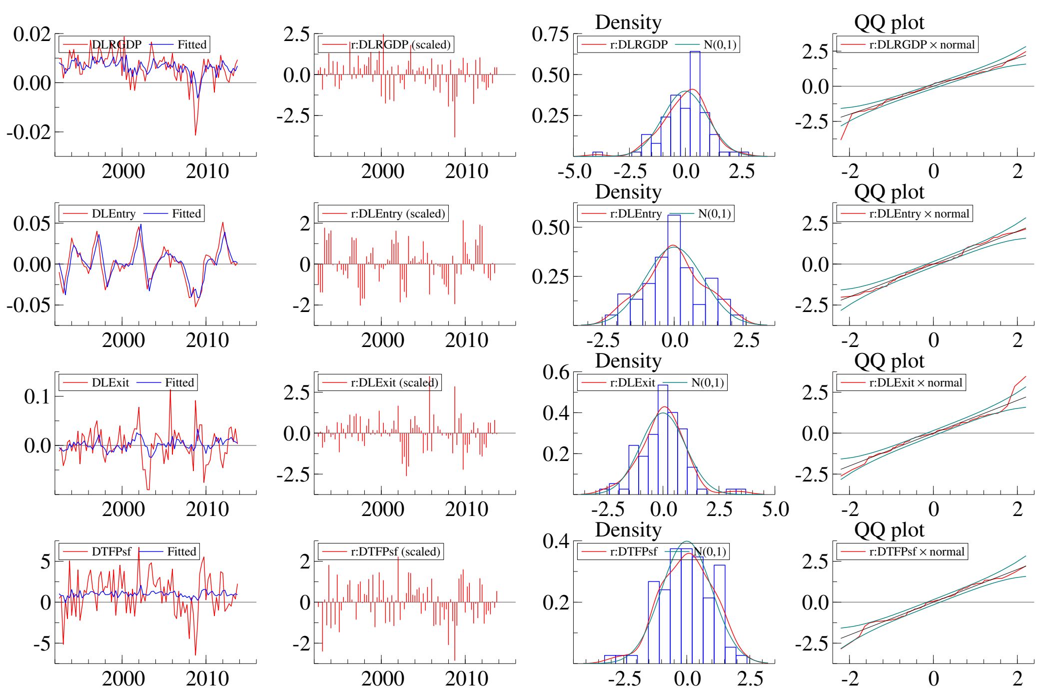


Figure 5: The estimated VAR model residuals for M5

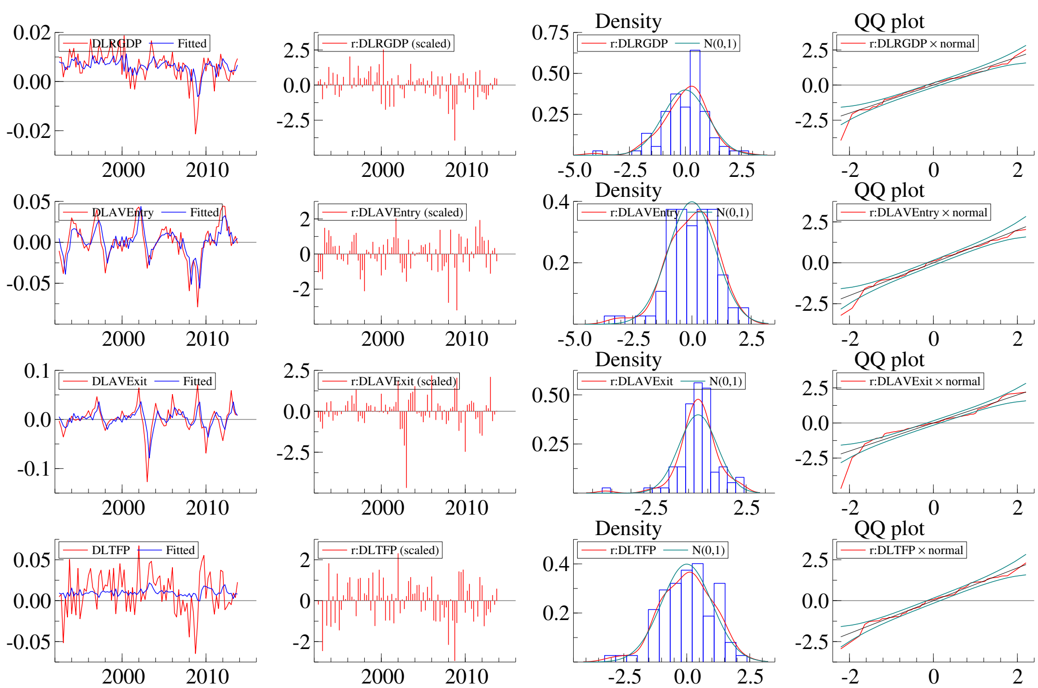
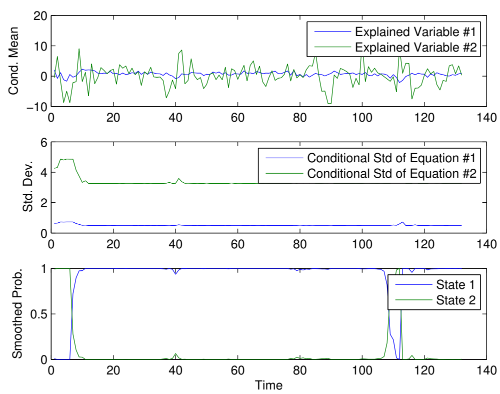
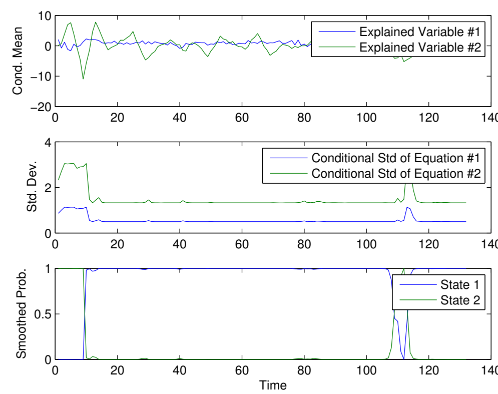


Figure 6: The MS-VAR models

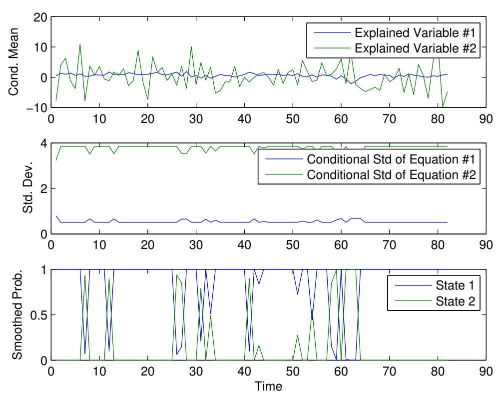
(a) M1a



(b) M2a



(c) M3a



(d) M4a

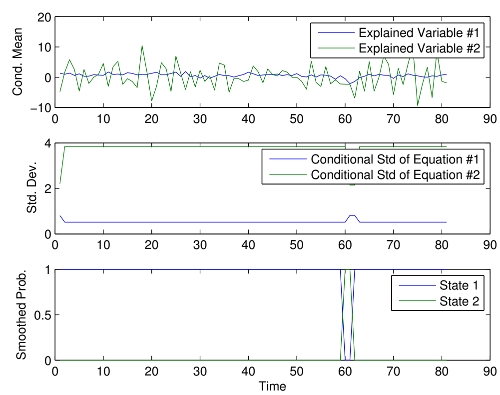


Figure 7: The M1b (Short-run restriction).

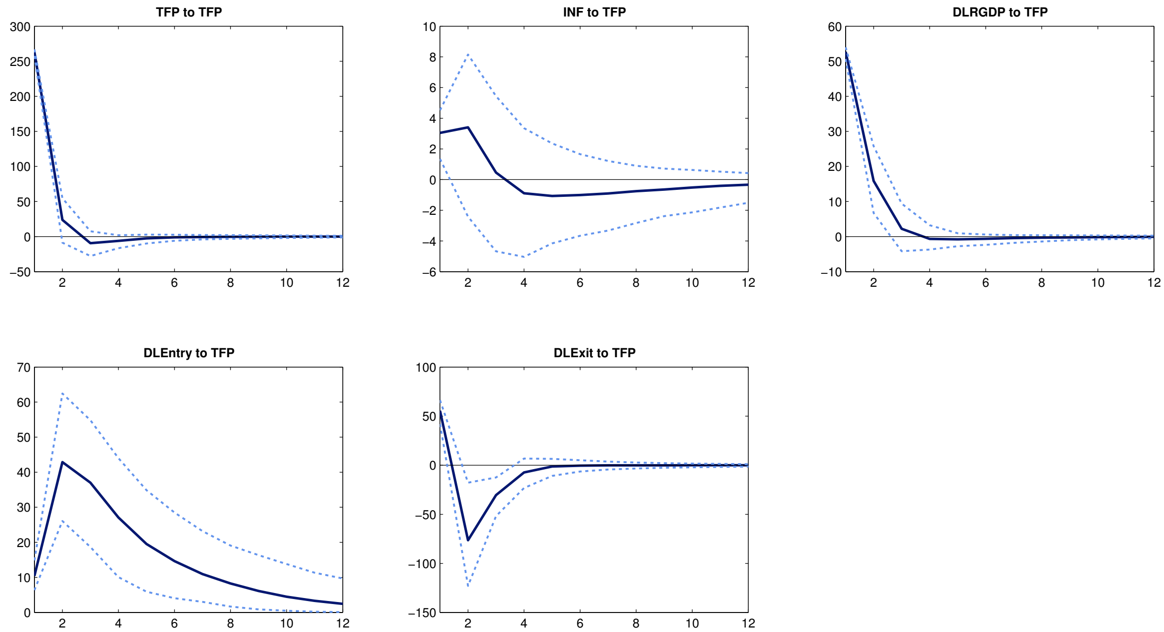


Figure 8: The M2b (Short-run restriction).

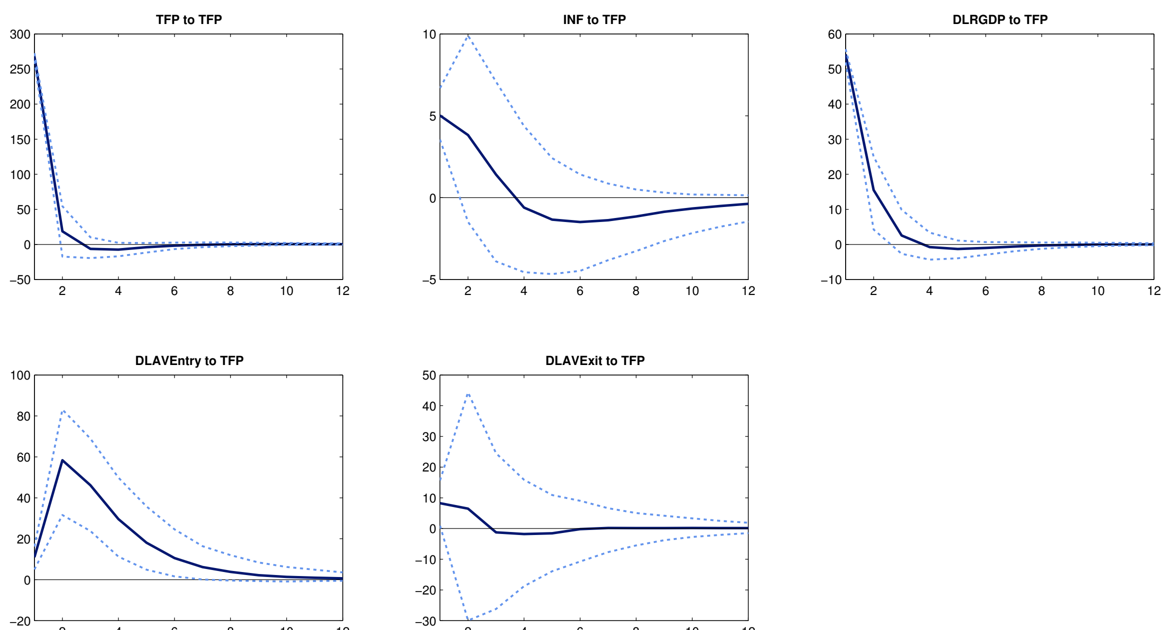


Figure 9: The M4b (Short-run restriction).

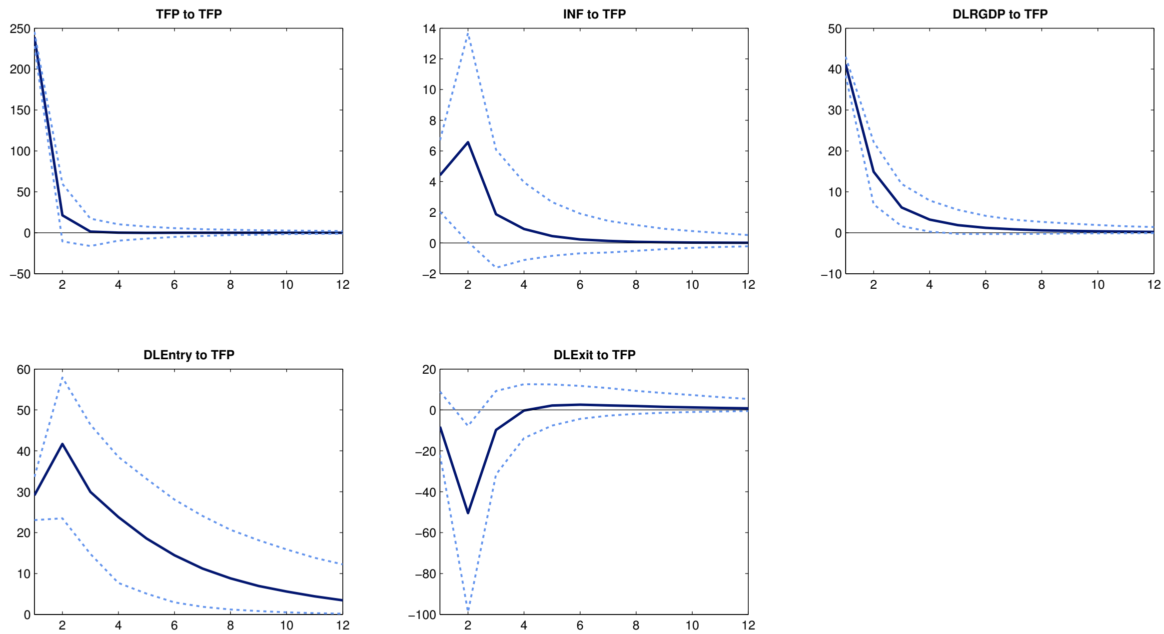


Figure 10: The M5b (Short-run restriction).

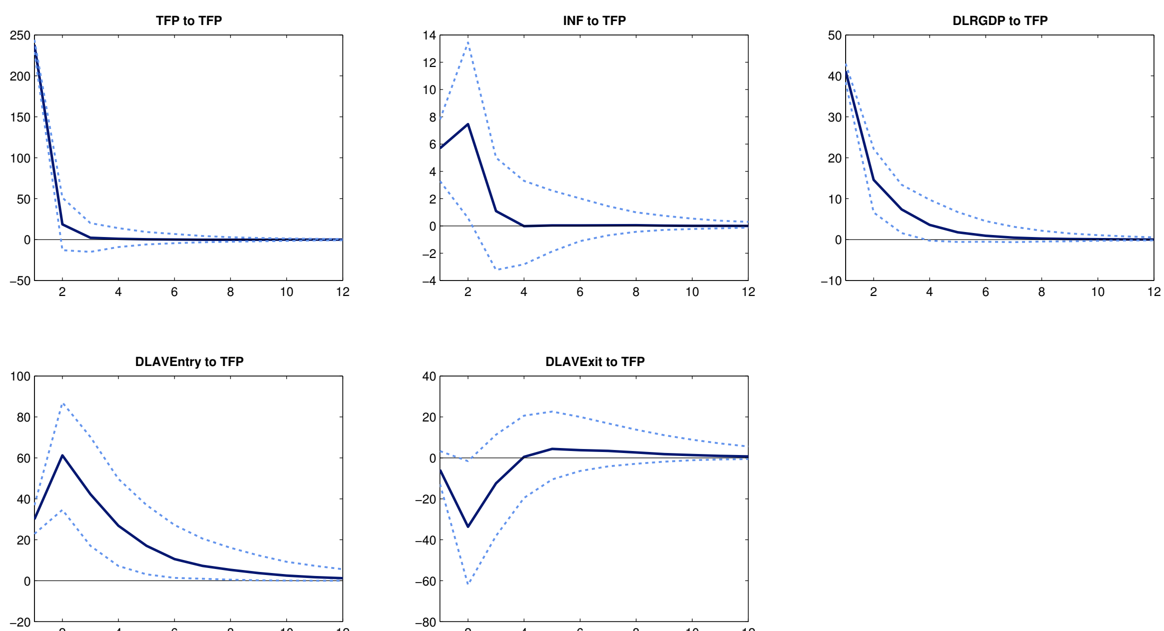


Figure 11: The R1 (Short-run restrictions).

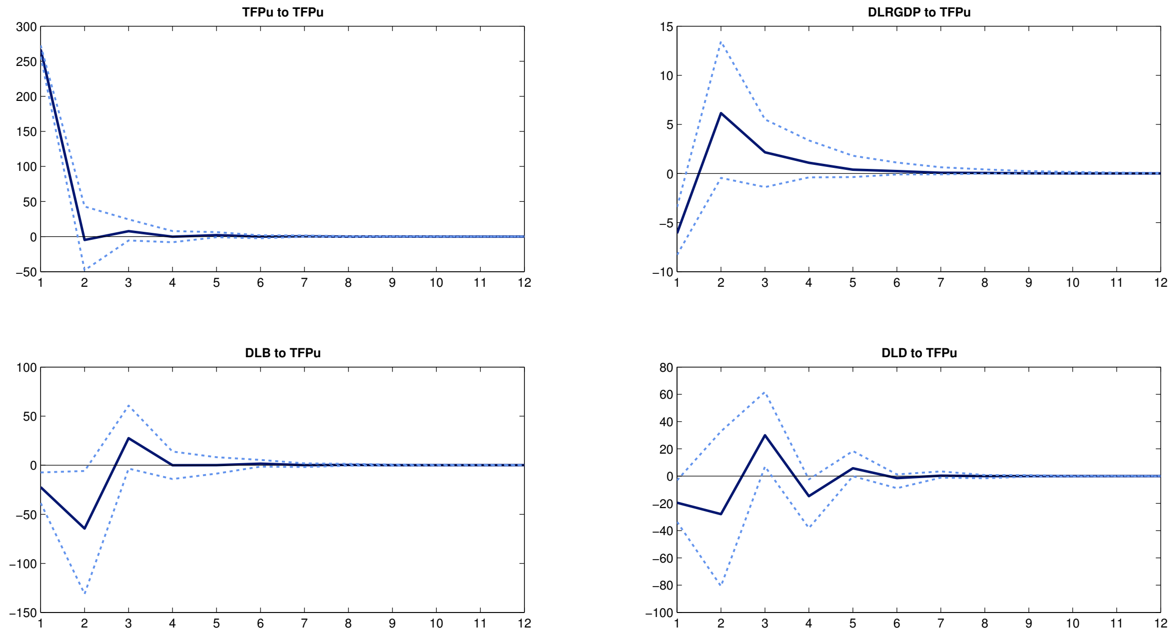


Figure 12: The R1b (Short-run restriction).

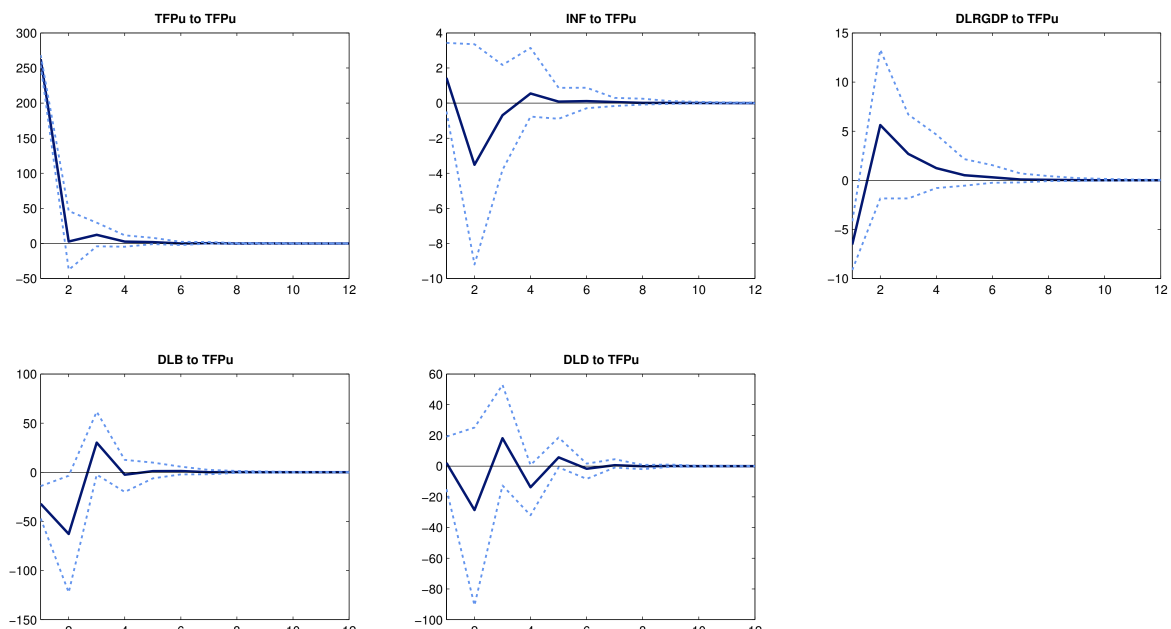


Figure 13: The R2 (Short-run restriction).

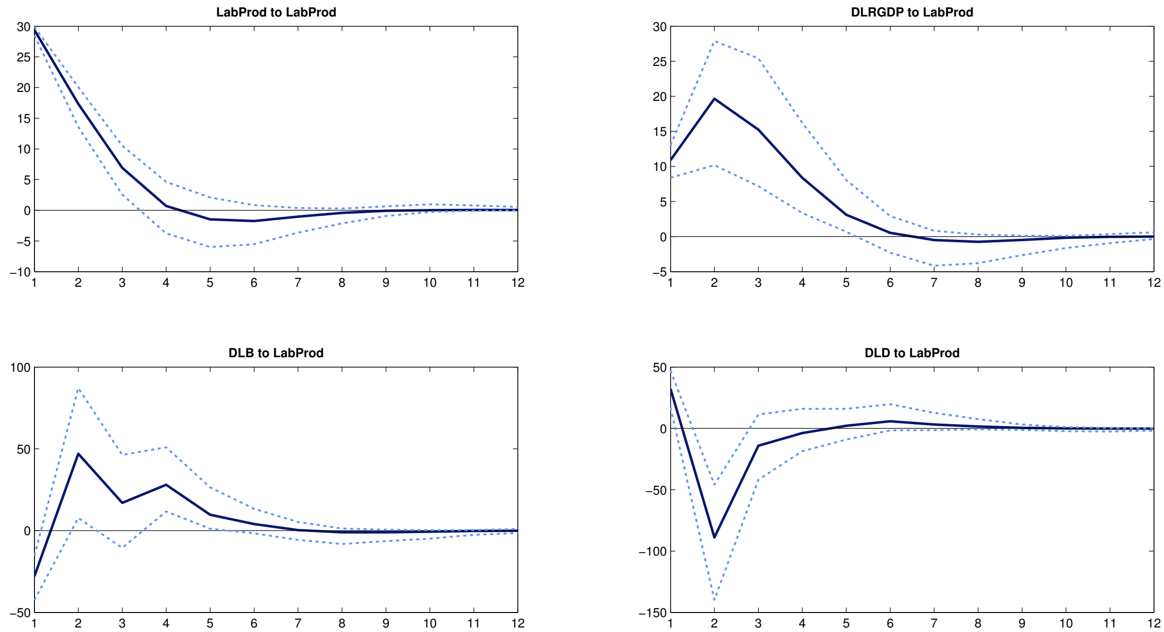


Figure 14: The R2b (Short-run restriction).

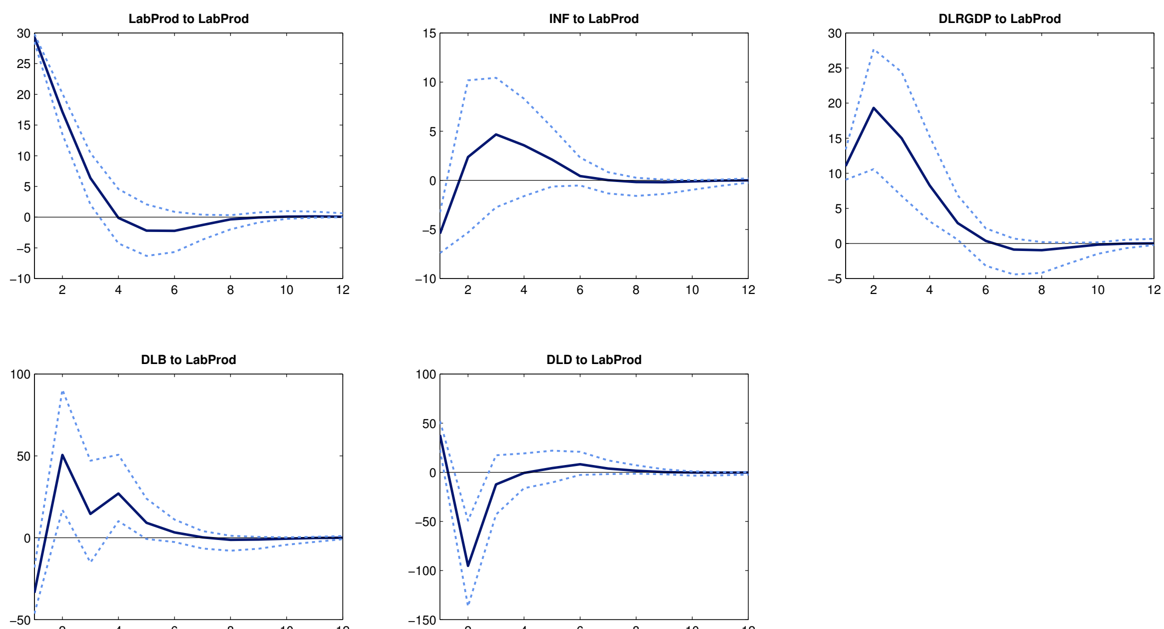


Figure 15: The R1 (Long-run restriction).

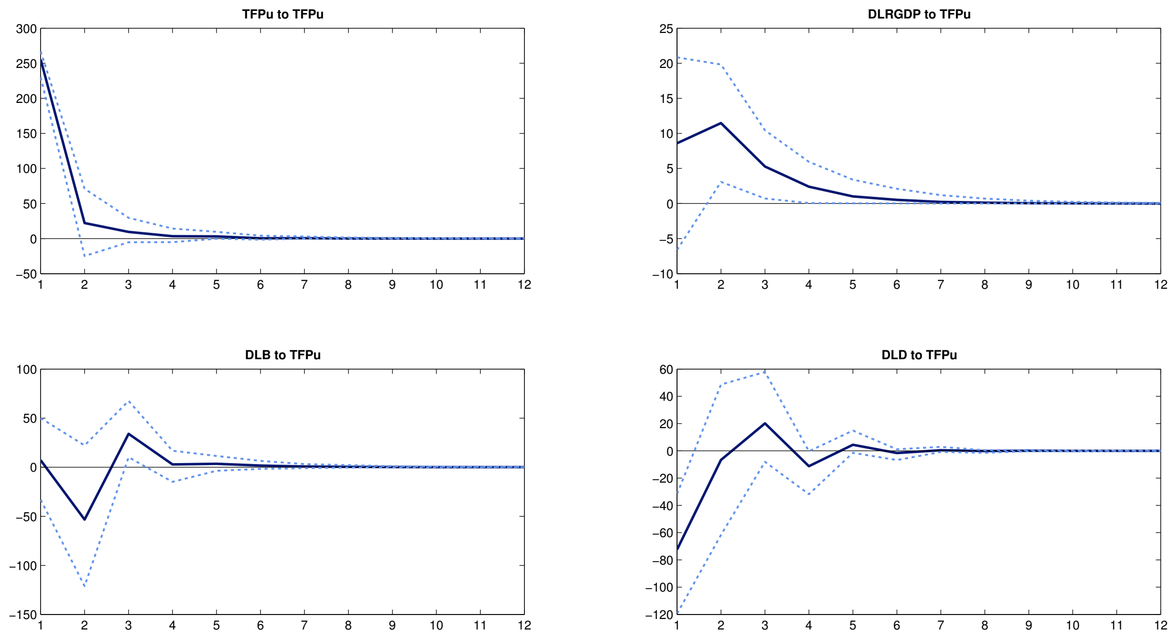


Figure 16: The R1b (Long-run restriction).

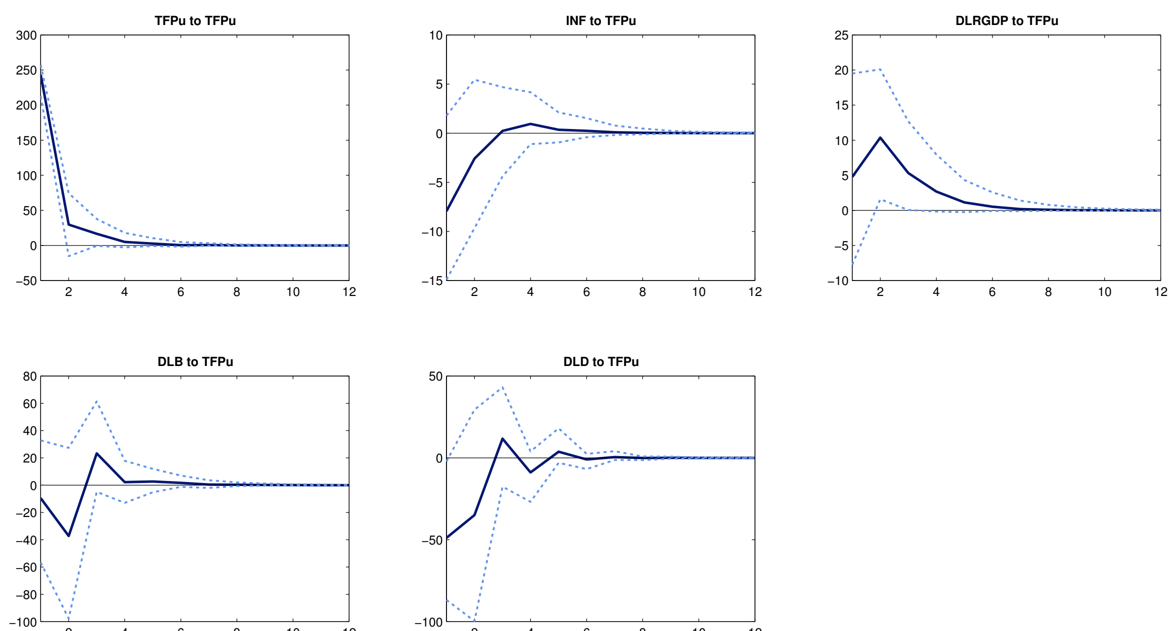


Figure 17: The R2 (Long-run restriction).

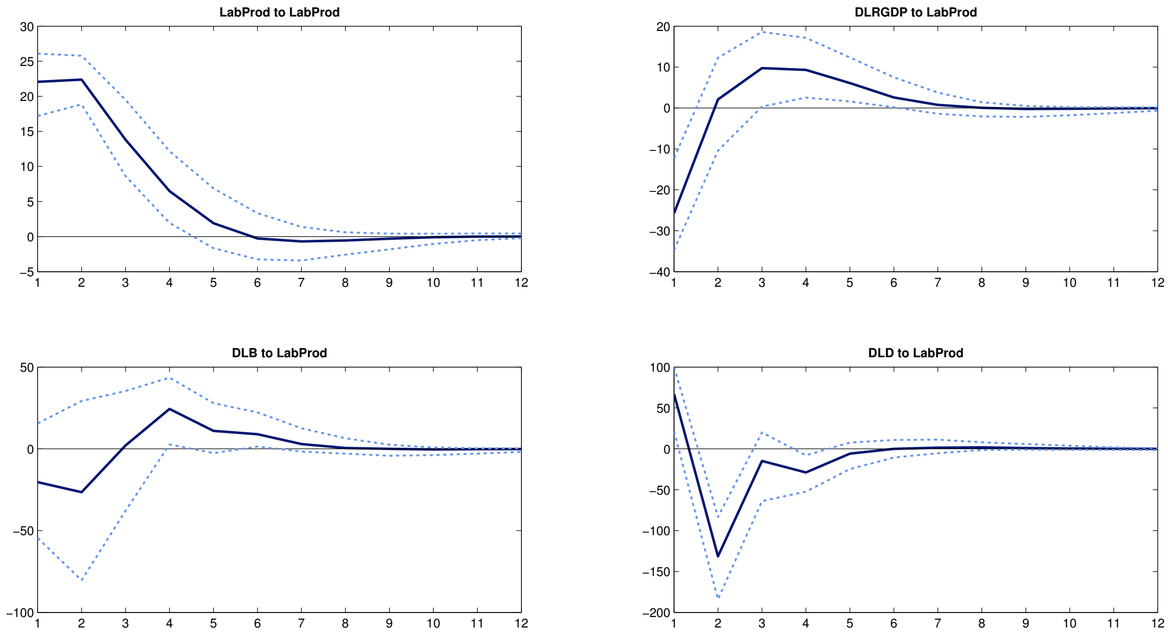


Figure 18: The R2b (Long-run restriction).

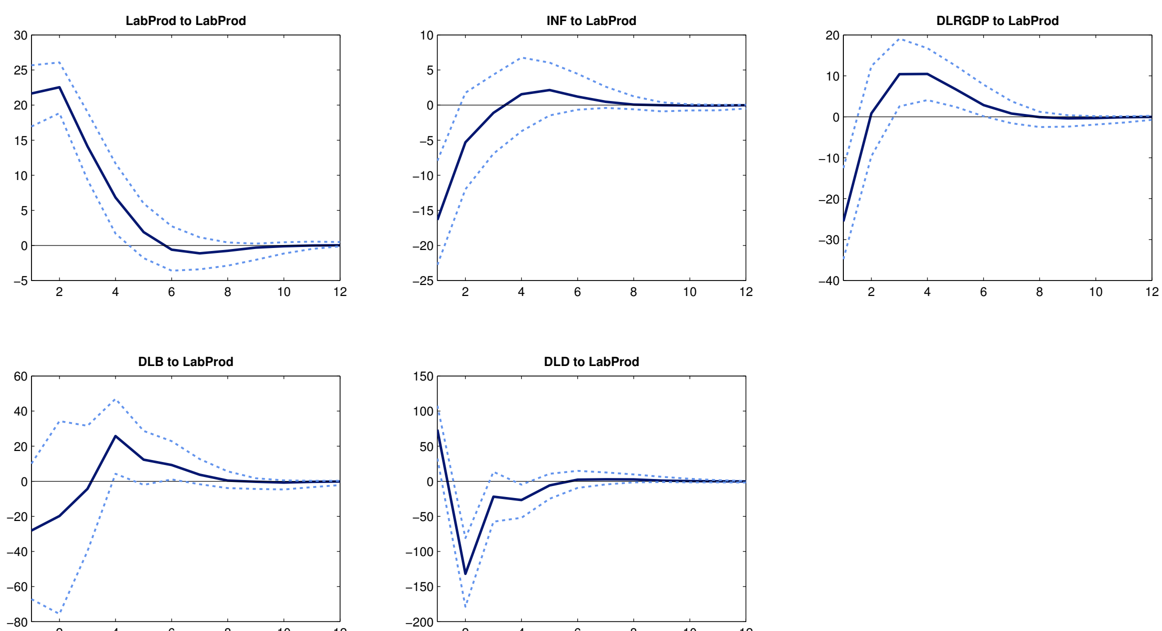


Figure 19: The M1 (Long-run restriction).

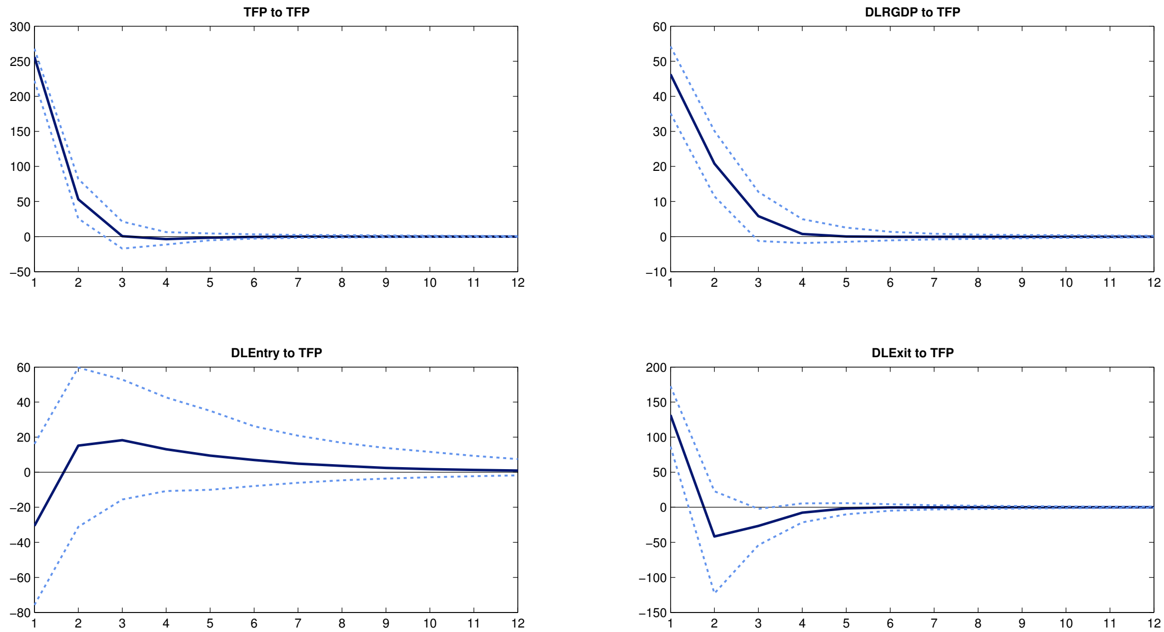


Figure 20: The M1b (Long-run restriction).

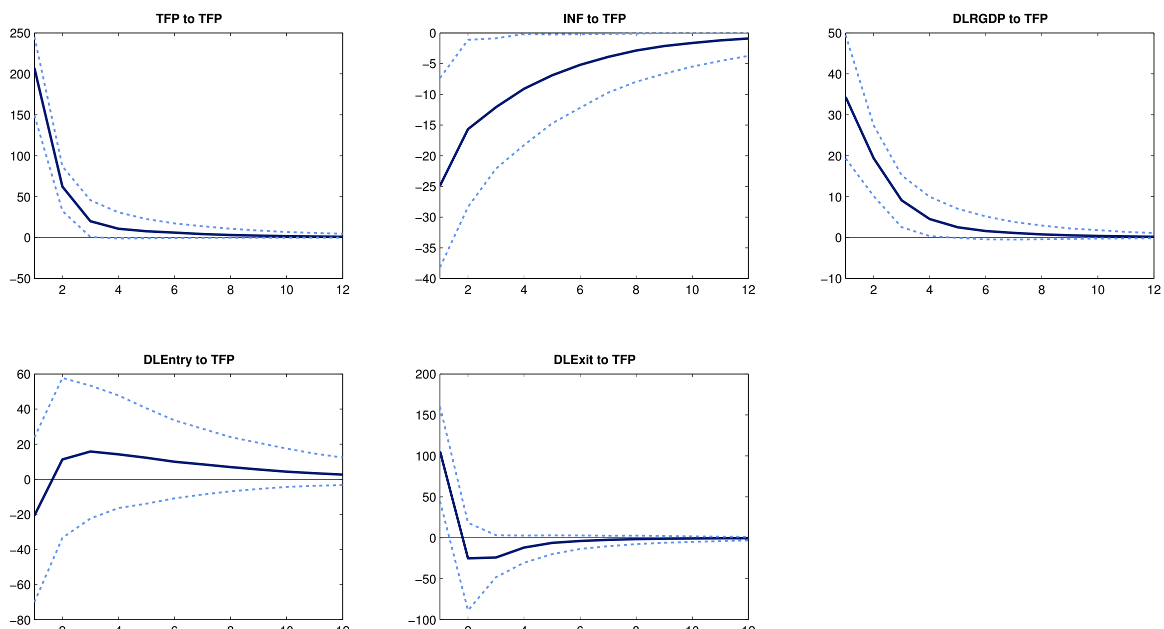


Figure 21: The M2 (Long-run restriction).

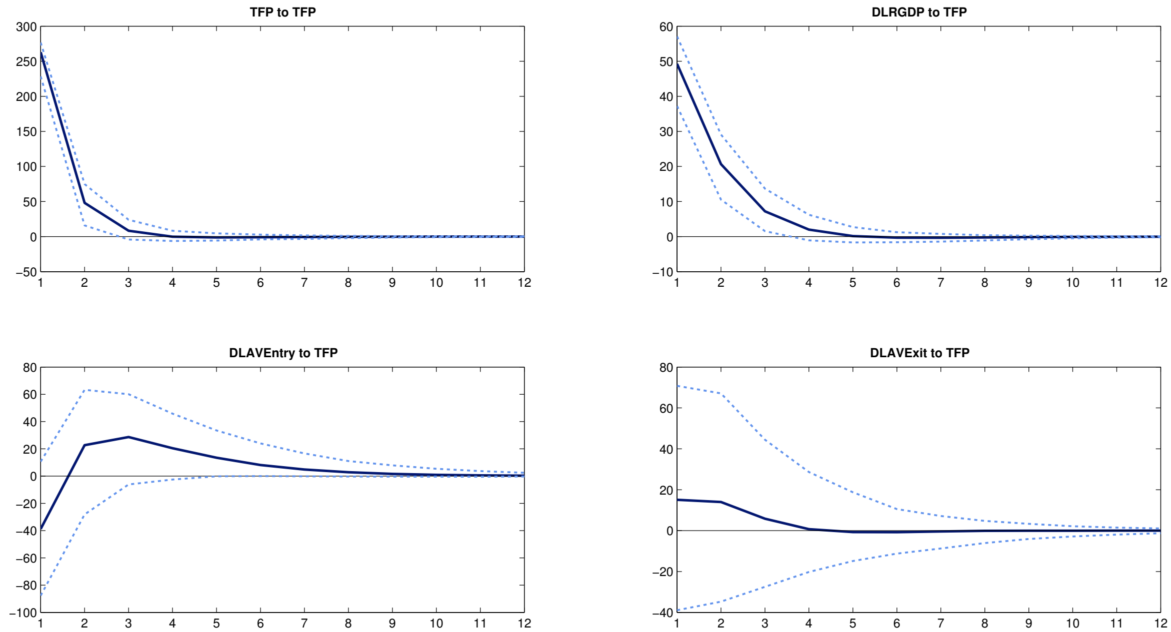


Figure 22: The M2b (Long-run restriction).

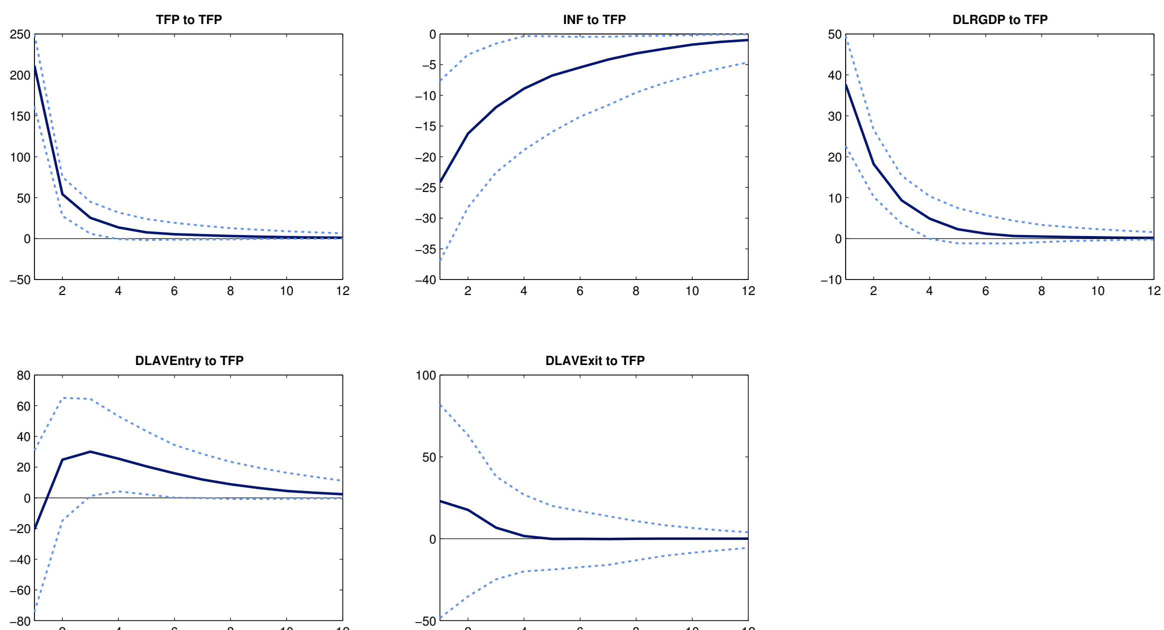


Figure 23: The M3 (Long-run restriction).

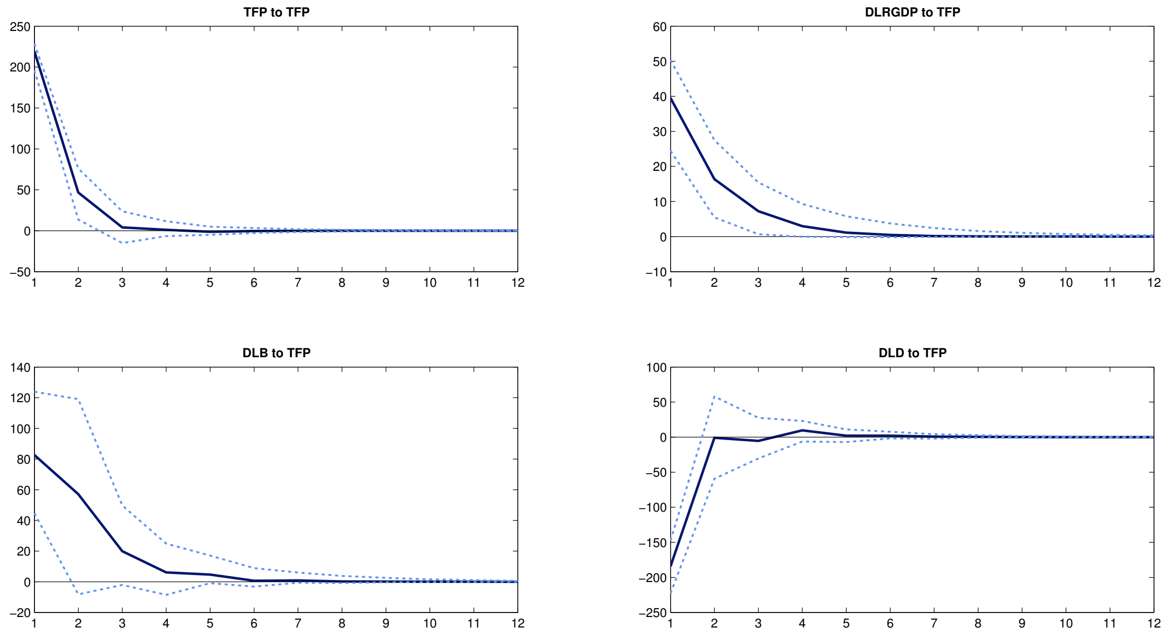


Figure 24: The M3b (Long-run restriction).

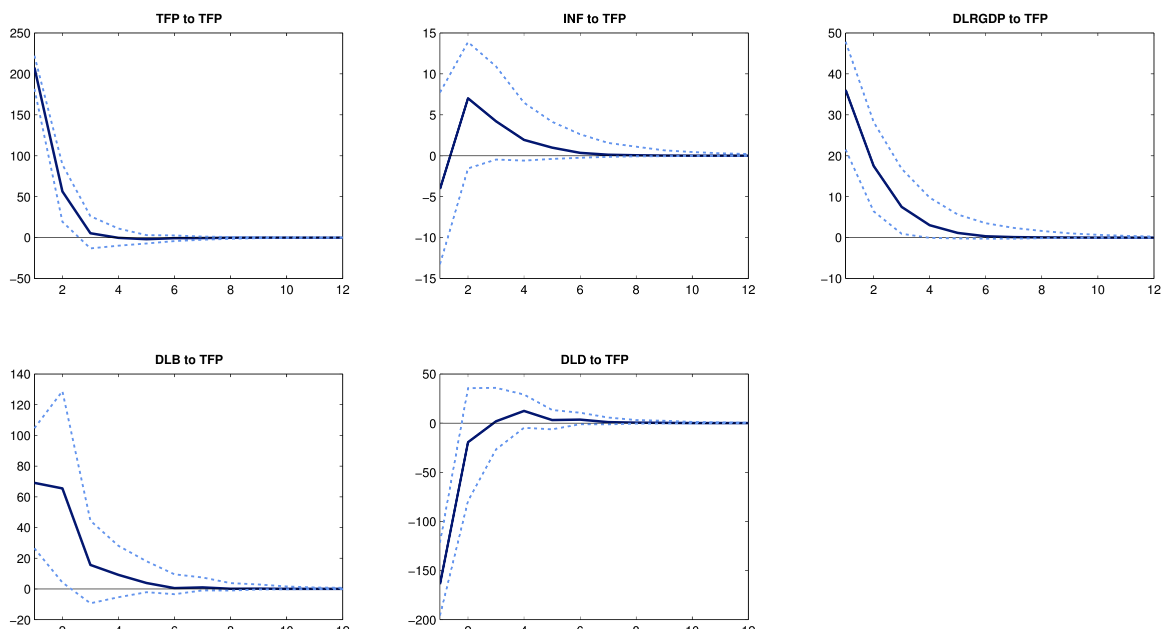


Figure 25: The M4 (Long-run restriction).

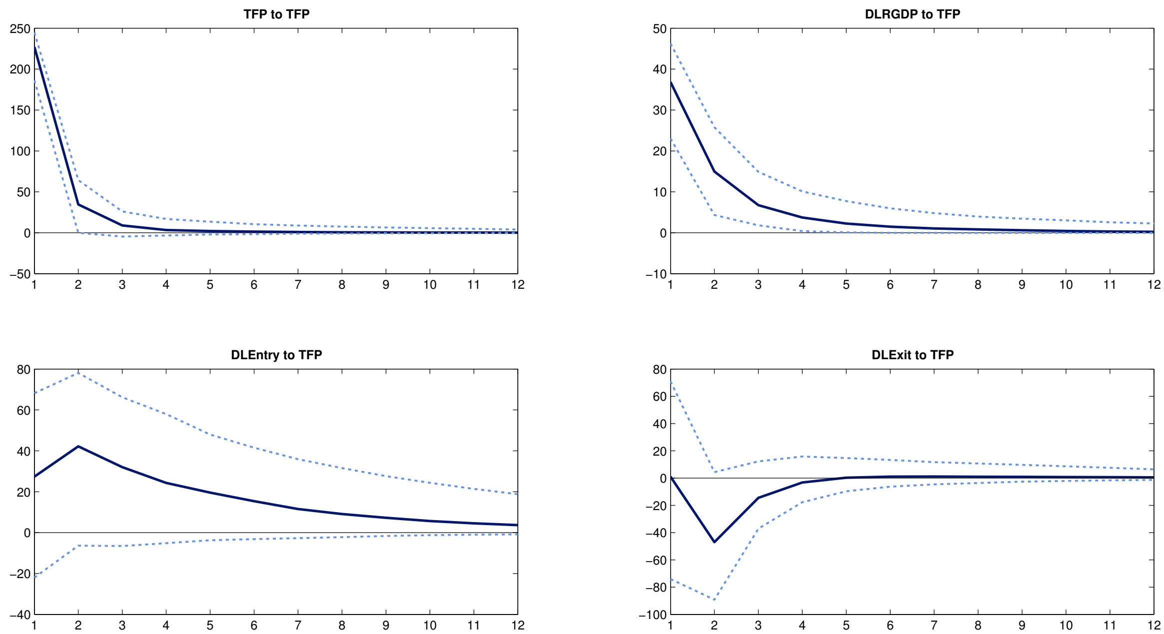


Figure 26: The M4b (Long-run restriction).

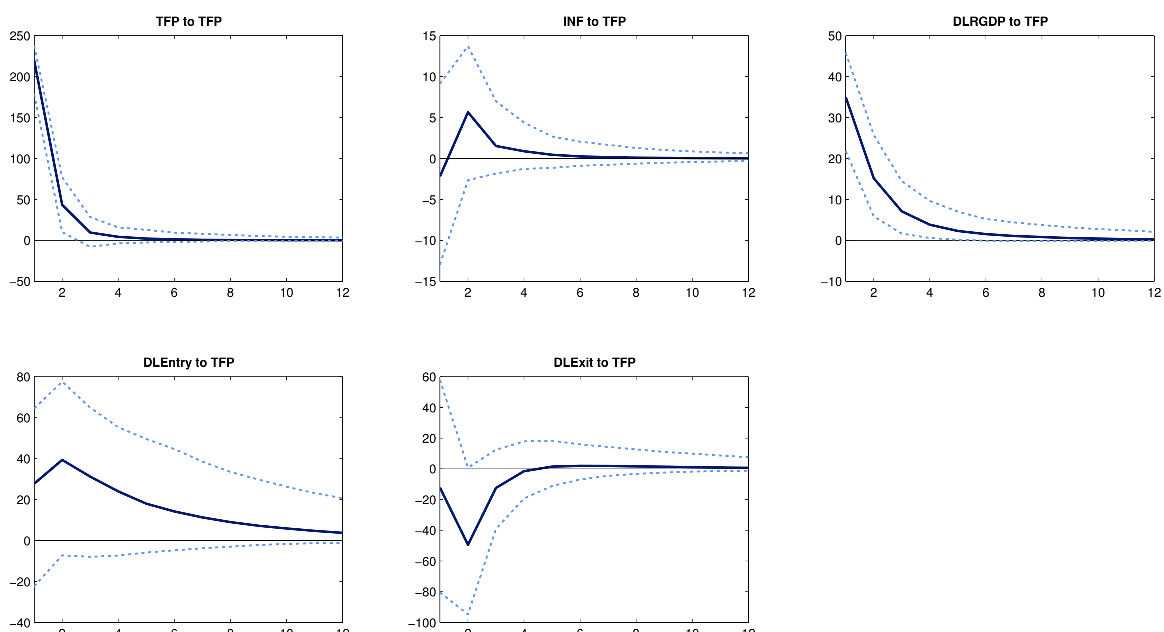


Figure 27: The M5 (Long-run restriction).

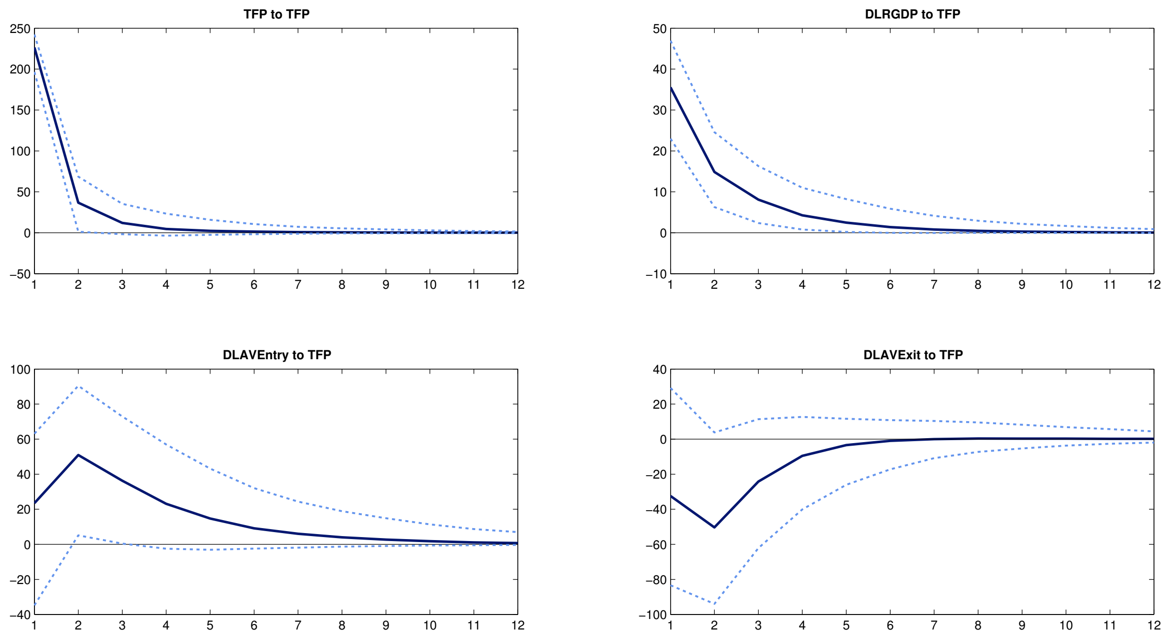


Figure 28: The M5b (Long-run restriction).