ISSN: 2281-1346

Department of Economics and Management

DEM Working Paper Series

Firms Dynamics and Business Cycle: New Disaggregated Data

Lorenza Rossi
(Università degli Studi di Pavia)

Emilio Zanetti Chini
(Università degli Studi di Pavia)

 # 151 (02-18)

Via San Felice, 5
I-27100 Pavia
economiaweb.unipv.it

February 2018

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: February 2018

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 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 pro-cyclical. According to a standard structural econometric analysis, exit overshoots its average level in the medium-run. Several robustness checks confirm these results, hence supporting the most recent theoretical literature.

JEL Codes: C13, C32, C40, E30, E32.
Keywords: Bayesian VAR, Entry, Exit, Productivity, State Space Models.
No. of words: 9,791 (6,791 in text + 3,000 in tables and figures)

---

*This paper has been supported by the Project 'MACFINROBODS' supported by the European Union. The Authors are grateful to Tommaso Proietti for having shared their 'TDS' package in Ox; this package has been properly modified by the Authors, who are responsible for any error or imprecision in the calculations. Moreover, the Authors thank Barbara Annicchiarico, Carolina Castagnetti, Efrem Castelnuovo, 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 possibly remain unfinished.

†Corresponding Author.

I Introduction

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

In this respect, any investigator could presume that there exists a plethora of dataset is built-up to conduct empirical analysis on this topic. This is only partially true: the dataset currently published by Census Bureau for the US economy is yearly and constitutes the only long-span data source on the argument; on the other side, the quarterly time series on firm's dynamics published by the Bureau of Labor Statistics (BLS, henceforth), start in 1993:Q2. Hence, the only time series suitable of immediate application in macroeconometric models faces with limited – and limiting – frequency. In facts, despite its rich variety of indicators at different level of aggregation, the yearly frequency of Census data limits the number of observations to use when estimating econometric models for business cycle; on the other hand, the starting date of BLS series do not allow a comprehensive long-span analysis as required by modern macroeconometric literature. This scarcity of macroeconomic data at higher frequency motivates this paper.

We rely on the literature on DSGE models for 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.[1] To the best of our knowledge few papers model firms' exit in a DSGE framework. Exceptions are Totzek (2009); Vilmi (2011); Cavallari (2015); Hamano and Zanetti (2017); Clementi and Palazzo (2016); Asturias et al. (2017); Rossi (2018). In particular, the latter author stands-out for her New-Keynesian DSGE model with endogenous entry, exit and banking sector. According to its dynamic properties in response to a productivity shock, firms' creation is found pro-cyclical; on the other hand, firms's exit is counter-cyclical on impact and, after few periods, overshoots its long-run value – taking almost seven years to return to its steady state level. This paper investigates the degree of conformity of these theoretical results – and particularly the overshooting of firms exit – with data which sample size is more proper for econometric analysis. Totzek (2009); Casares and Poutinau (2014); Hamano and Zanetti (2017) also provide a model with endogenous exit. However, they consider a perfect financial market and do not investigate on the overshooting of firms destruction. Also they do not provide an empirical analysis of firms dynamics in terms of impulse response functions in face of a productivity shock.

Our contribution is twofold. First, we make available for economic analysis new disaggregated series on firms dynamics in the US, which can be used as an alternative and longer source than the BLS quarterly data. Second, we carry out a business cycle analysis of the disaggregated series, jointly with a structural VAR analysis to test their dynamic performances in response to a positive productivity shock. With respect to the first contribution, we disaggregate the yearly series provided by the Census Bureau[2] We consider two different types of 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 ordinary least-square (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), hence allowing generality and flexibility besides maintaining statistical robustness.

We anticipate that the UCM-based series are superior in terms of statistical accuracy with respect to the ones resulting by applying other naive models. 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. Both the UC models rely on univariate estimations and consider a single regressor variable, here represented by the industrial production. Thus, the dynamics of a single indicator might influence the final estimates of disaggregated process more than expected. In order to circumvent this problem, we repeat the same univariate UCM on all the indicators of a new, large macroeconomic dataset and average the resulting univariate UC-based series. This ad-hoc combination produces new series which take in account of all the information on the US economy.[3]

After a comprehensive discussion of data and methods, respectively in Sections II and III, Section IV extracts the trend and cycle components via standard first difference operator, the Hodrick and Prescott (1997) (HP) and Baxter and King (1999) (BK) filters and, subsequently, classifies them as leading or lagging indicators of the business cycle by looking at the maximum absolute value of cross-correlations between the cyclical components of ENTRY and EXIT and the one of GDP. According to our results, while ENTRY is lagging and pro-cyclical and EXIT is a leading and countercyclical indicator.

A structural econometric analysis is discussed in Section V, where a battery of bayesian vector autoregressive (BVAR) models constituted by various alternative proxies of ENTRY and EXIT (that is, the two disaggregated series versus the BLS data), real GDP and inflation is used to investigate the dynamics of impulse response function (IRF) to a productivity shock. For all the models here adopted, firms creation is found persistent, while firms destruction is overshooting its long-run level for all proxies used; as a consequence of the order imposed to the system of equations, the real GDP increases and inflation declines. These results are compatible with recent macroeconomic literature on endogenous firms dynamics and are confirmed by several robustness checks.[4]

Finally, Section VI concludes. An associated Supplement reports additional empirical evidence mentioned in the text, jointly with mathematical details on statistical methods and the robustness checks.

II Data

The Business Dynamics Statistics (BDS), published by Census Bureau Research Data Centers, is only official long-span dataset for U.S economy. It gives information about total number of firms, establishments and workers, establishments opened and closed, job creation and destruction and other 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.[5]

Analysts interested in higher frequency samples can rely on two other sources:

(i) 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 with starting date in 1993:Q2. These series are direct observations of the number of firms/establishments.

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

Despite their similar nature and 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. A practical implication of this remark is that no interpolation is possible between the BLS and 'Economagic' series, so 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, 'Economagic' data are the basis of a recent strand of literature despite the exact definition of incorporation used is not available.[6]

Thus, 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 (source: FRED, Federal Reserve Bank of St. Louis). the estimated bars are largely below the critical values of ± 0.20, corresponding to the blue bars; see Figure 1 of Supplement. Our contribution is precisely in the production of new time series on firm's dynamics capable to overcome this measurement error problem.

As the next Section explains, 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. To this aim, we use 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).[7] Namely, the yearly ENTRY (EXIT) series are regressed on all these variables in quarterly frequency by in order to obtain a new quarterly series of the regressand. This solution conveys new time series so similar to the one obtained via univariate regression that further methodological investigation seems us purely academic.

III 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 (or '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: the first one is a UC representation of the Chow-Lin regression model (CL-UCM); the second one is a more general autoregressive distributed lag (ADL-

UCM) model. Both of them are estimated by AKF[8]. 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 $t = 1, \ldots, T$; multiple time series are denoted in bold and matrices in capital.

The Chow-Lin 'Naive' Method

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.

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[9]. 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).

Combining multiple disaggregates

The UCM disaggregation previously described holds for a single time series, that is, In terms of system (2), 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 $\hat{\mathbf{Y}}_\mathbf{t} = [\hat{\mathbf{y}}_\mathbf{1}, \ldots, \hat{\mathbf{y}}_\mathbf{n}, \ldots, \hat{\mathbf{y}}_\mathbf{N}]$, with y_n representing the $[T \times 1]$ vector of the process disaggregated via system (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 y_t in (2).

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

$$\hat{y_t}^C \doteq \frac{1}{N}\sum_{n=1}^{N} y_{t,n}^s, \tag{7}$$

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 (7) conveys the series labeled $AVEntry$ and $AVExit$ – to underline the fact that they are an average of many single processes.

IV Descriptive Analysis

A graphical inspection of the series estimated from all the three models introduced in previous section is shown in Figure 1. 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 conspicuous spread between correlations of firms' creation and destruction 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 rejects the null hypothesis of one cointegrating relation between the series of ENTRY and IP – properly aggregated – strongly.[10] 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 III.

The output of CL-UCM specification are characterized by absence of outliers and noise. Differently and 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 reduction in entry is more than the double, in absolute value, under recession. The quasi-noisy behavior 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; see also the graphical comparison in Figure 2 in Supplement.

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 1 of Supplement, 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. We consider the LR-test also as a valuable criterion to select the model specification to use for each of the 134 indicators of FRED-MD. Once these has been selected, the combination of the disaggregated series maintaining same UCM is done ex-post. In this way we obtain series alternative but qualitatively very similar to the ones derived from a single indicator previously analyzed; see Figure 3 in Supplement. In principle, this can be explained from the simple fact that many of the variables in that dataset have mutually different paths. This compensates many of the (potential) differences that one could notice when the disaggregation is done by using the same indicators singularly. Thus, a simple combination of many UCM-based outcomes with no economically meaningful selection of the indicators should be treated with caution.

The trend and cyclical components of our new series are displayed in Figure 2: all the series have the same path in both the components, and just in the trend some difference is visible between the two filters used; the cyclical components of univariate and Combination UCM are substantially coincident. A different scenario, plotted in Figure 3, is obtained if repeating the same graphical check on BLS data: the BK filter conveys a cycle considerably smoother 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 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 4 of Supplement: our series anticipate the recession phase in ENTRY, albeit BLS data are 'leading' in EXIT measures.

At the light of what above described, we assert that the nature of the BLS proxies is different from our disaggregated ones. Hence, all of them have to be classified 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 and 8 quarters for the EXIT[11]. According to Table 3 we find that:

(i) the BDS disaggregated ENTRY series 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. This constitutes the only difference with AVEntry.

(ii) On the contrary, AVExit 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-filtered cycle is counter-cyclical and leading; instead, the opposite holds for the HP-filtered cycle (pro-cyclical with two years of lagging).

(iii) The contemporaneous correlations with RGDP are positive for ENTRY and negative for EXIT, but these 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).

V Structural Analysis

Once the statistical properties of the new disaggregated series have been investigated, it is possible to use them for structural econometric analysis. To this aim, we estimate small BVAR models and show the IRFs to orthogonalized shocks to labor productivity as well as to the total factor productivity (TFP, henceforth); see the Supplement for IRFs associated to TFP shocks.[12] We use data on RGDP, inflation, and the three alternative proxies of firms' dynamics, and namely:

(i) the disaggregated series of firms AVEntry and AVExit obtained with the multivariate UC method;

(ii) the series of firms ENTRY and EXIT obtained with the univariate UC method;

(iii) the BLS quarterly series of firms BIRTHS and DEATHS.

The bayesian approach to structural analysis has been considered for its versatility and capability to address the issue of short sample size. In fact, it avoids sampling errors in estimation of IRF bands that may occur when sample is short or, equivalently, when the model is highly over parameterized, see Sims and Zha (1998). We estimate five different BVAR specifications, which corresponds to five different alternative Entry and Exit proxies (labeled M1–M5, henceforth), see Table 2 for model definitions and description of the samples. Namely, we consider a sample size 1993:Q2-2013:Q4 for the BVAR estimated with the BLS series of BIRTHS and DEATHS, while a longer sample (1983:Q1-2013:Q4) is instead used for the BVAR estimated using the disaggregated (AV)Entry and (AV)Exit. In this way, we exclude the period of high inflation and the monetary regime change of the early-80s and consider only the Volcker-Greenspan-Bernanke period. Two BVAR using the disaggregated series for the sample 1993Q2-2013Q4 are also considered to compare their analogues which adopt BLS data. All the variables under consideration are expressed in logarithms of the levels. Precisely, for each variable Y we adopt LY to label its log-transform[13]. As it is now standard in the literature, this implicitly allows for the possible presence of cointegrating relations, without imposing restrictions on the long-run properties of the model, see Sims et al. (1990). The reduced for of the BVAR is:

$$\boldsymbol{y_t} = \boldsymbol{A_1 y_{t-1}} + \boldsymbol{A_2 y_{t-2}} + ... + \boldsymbol{A_p y_{t-p}} + \boldsymbol{u_t}, \tag{8}$$

where $\boldsymbol{A_j}, (j = 1 \dots p)$ are $(N \times N)$ coefficient matrices and $\boldsymbol{u_t} = [u_{LabProd,t,}, u_{RGDP,t}, u_{ENTRY,t}, u_{EXIT,t}]'$ is a white noise vector of time series with $\boldsymbol{u_t} \sim (\mathbf{0}, \boldsymbol{\Sigma_u})$.

For each system under investigation, we estimate a BVAR(2) using Minnesota Priors of 0.8 on the autoregressive coefficient of the first lag for the distribution of parameters. These priors are justified by the short sample size. In fact, in this framework, it is assumed that the VAR residual variance-covariance matrix $\boldsymbol{\Sigma_u}$ is known; here we make use the entire variance-covariance matrix of the VAR system estimated by OLS[14]. As a robustness check, we estimate the same model specifications under Normal Diffuse Priors; a prior of 0.8 on the autoregressive coefficient of the first lag has been set also in this case; the results are reported in the Supplement. Finally, we choose a lower triangular Cholesky identification and order LLabProd first, such that on impact shocks to productivity affect the other variables, while shocks to the other variables do not affect productivity on impact. This ordering has been widely used in the literature, see Bloom (2009) *inter alia*. In what follows, we consider responses to 1 percent shock to the labor productivity.

The IRFs resulting from the BVAR under Minnesota Priors are plotted in Figures 4 – 8.[15] The real GDP increases in response to a productivity shock, while inflation falls down; this is exactly the picture depicted by New Keynesian theory. According to all the specification here considered, firms' creation is pro-cyclical and persistent for all the proxies considered in the five different specifications of the BVAR model. In addition, firms destruction is countercyclical and, remarkably, overshoots it long run level in the medium run, as suggested by recent the theoretical model described in Rossi (2018). These results are robust to the use of the Normal Diffuse Priors: all IRFs shown in Figures 5 – 9 of the Supplement have the same patterns of the ones obtained with Minnesota priors, despite the confidence interval enlarges, as expected by standard econometric literature. This is particularly evident for the disaggregated series of EXIT.

These econometric models confirm the descriptive findings reported in Table 3, hence suggesting that the series of entry and exit are, respectively, pro-cyclical and countercyclical if conditioned to a productivity shock. We underline that the other different proxies of firms destructions seem to be less inertial than firms creations and overshoot their long run level in the medium run. Consequently, our disaggregated, considerably longer series are an interesting alternative to the use of the BLS ones in the empirical analysis.

Figures 10 – 14 of the Supplement displays the IRF resulting from same exercise if assuming that LabProd, ENTRY and EXIT in their different proxies, are in log-differences; as a consequence, in this sub-groups of models, the variable of inflation is replaced by CPI. The use of data in log-deviations corresponding to M1b – M5b has effects on the magnitude of the impact but not in the general path of the whole responses. In some cases, the the response is radically different, as in the case of M2b, where Exit seems to increase to a positive shock of the labor productivity and in M4b and M5b whereas the effects on Exit is not clear due to large parameter uncertainty; see figure 11 of Supplement).

Figures 15 – 19 of the Supplement plots the M1–M5 where the labor productivity is repleced by an alternative measure, the Total Factor Productivity (TFP) according to the definition and data by Basu et al. (2006). Still, a BVAR(1) with Minnesota priors with 0.8 on the autoregressive coefficient of the first lag a lower triangular Cholesky identification, ordering the TFP first – such that on impact shocks to the productivity index affect the other variables, while shocks to the other variables do not affect the productivity index on impact – are assumed. Overall, the results reported in Figures 4 – 8 are confirmed.

VI Conclusions

The recent advances in macroeconomic theory make the availability of new time series data on firms' dynamics an important 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 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 pro-cyclical. Moreover, a productivity shock is associated to a positive and persistent response of entry and a negative and persistent response of exit.

Disaggregation methods should account for the increase in dimension and availability of macroeconomic datasets and possible nonlinearity of the indicator(s). This suggests that two important topics for future methodological research are the adoption and/or adaptation of variable selection algorithms recently appeared in econometrics and the introduction of a high-dimention nonlinear structure in the state-space representation for disaggregation.

Notes

1 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.

2 All the 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.

3 We remark that averaging many different disaggregated processes assumes implicitly that each indicator has equal weight in defining the combined time series. This choice could be not the best, so that we will consider this strategy only as robustness check. More refined methods that use factor analysis like the contributions by see Frale et al. (2011) and Grassi et al. (2015) are available and constitute a possible development in this sense.

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

5 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.

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

7 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.

8 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.

9 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.

10 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 it corresponds to test the null hypothesis of no cointegration versus the alternative of 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.

11 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 2 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.

12 Here, we consider labor productivity for two reasons. First, because it is the variable considered in the theoretical model by Bilbiie et al. (2012); Colciago and Rossi (2015); Hamano and Zanetti (2017); Rossi (2018), among others. Secondly, it is available in levels and in quarterly terms from FRED database, while TFP is available only in annual terms. Nevertheless, for robust check we use the standard TFP measure computed by Basu et al. (2006). This series is quarterly, but available only in log-difference terms. For this reason we consider the TFP only in the BVAR estimated in log-deviations. In the Technical Appendix,these estimations will be compared with the same BVAR where also labor productivity is in logarithm difference.

13 For example, in M1 the N vector of observable time series is $\boldsymbol{y_t} = [LLabProd_t, LCPI_t, LRGDP_t, LAVENTRY_t, LAVEXIT_t]'$, where LLabProd is the logarithm of labor productivity, $LCPI_t$ is the logarithm of Core CPI Index, while $LRGDP_t$, the logarithm of real GDP, LAVEntry and LAVExit are the logarithms of the disaggregated series of AVEntry and AVExit (that is the ones obtained using the multivariate UC method) and the vector length is T with $T = 122$ at the full sample. All these series has been dowloaded from FRED database.

14 Since the model estimates all the equations simultaneously, the assumption of a diagonal matrix of the original Minnesota prior, which in turn implies independence between the VAR coefficients of different equations and justified at that time of limited computational power as it estimates the model equation by equation, is not necessary.

15 In all of the plots, the median responses to one-standard-deviation increase in the innovations to labor productivity are depicted by solid lines, while dashed lines represent 84 and 16 % confidence intervals.

References

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

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

Asturias, J., S. Hur, T. Kehoe, and K. Ruhl (2017): "Firm Entry and Exit and Aggregate Growth," NBER Working Paper No. 23202.

Basu, S., J. Fernald, and M. Kimball (2006): "Are technology improvements contractionary," *American Economic Review*, 1418–1448.

Baxter, M. and R. King (1999): "Measuring business cycles: approximate band-pass filters for economic time series," *The Review of Economics and Statistics*, 81, 575–593.

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

Bergin, P., L. Feng, and C. Lin (2014): "Financial Frictions and Firms Dynamics," NBER Working Paper no 20099.

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

Bloom, N. (2009): "The impact of uncertainty shocks," *Econometrica*, 77, 623–685.

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

Cavallari, L. (2015): "Entry costs and the dynamics of business formation," *Journal of Macroeconomics*, 44 (C), 312–326.

Chow, G. and A. Lin (1971): "Best Linear Unbiased Interpolation, distribution and extrapolation of time series by related series," *The review of Economics and Statistics*, 53, 372–375.

Clementi, G. and B. Palazzo (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomic*, 8, 1–41.

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

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

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

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

Frale, C., M. Marcellino, G. Mazzi, and T. Proietti (2011): "EUROMIND: a Monthly Indicator of Euro Area Economic Conditions," *Journal of Royal Statistical Association, ser. A*, 174, 439–470.

Grassi, S., T. Proietti, C. Frale, M. Marcellino, and G. Mazzi (2015): "EuroMInd-C: Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries," *International Journal of Forecasting*, 31, 712–732.

Hamano, M. and F. Zanetti (2017): "Endogenous Product Turnover and Macroeconomic Dynamics," *Review of Economic Dynamics*, 26, 263–279.

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

Hodrick, R. and E. Prescott (1997): "Postwar US business cycles: an empirical investigation," *Journal of Money, Credit, and Banking*, 1–16.

Jaimovich, N. and M. Floetotto (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 L. Rossi (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.

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

McCracken, M. and S. Ng (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.

Rossi, L. (2018): "The Overshooting of Firms Destruction, Banks and Productivity Shocks," DEM Working Paper No.147, University of Pavia.

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

SIMS, C., J. STOCK, AND M. WATSON (1990): "Inference in Linear Time Series Models with some Unit Roots," *Econometrica*, 58, 113–144.

SIMS, C. AND T. ZHA (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949–968.

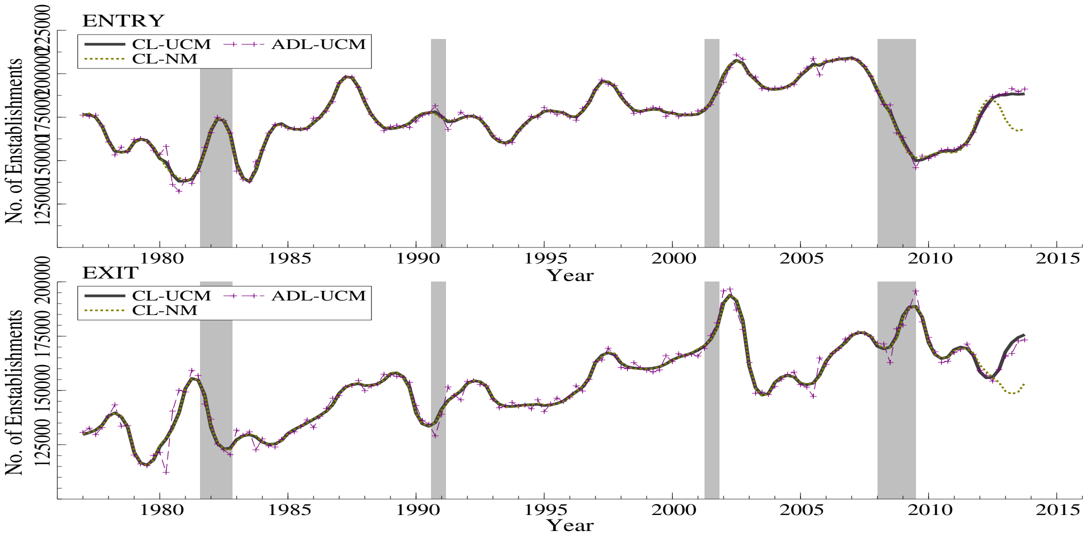
TOTZEK, A. (2009): "Firms' Heterogeneity, Endogenous Entry, and Exit Decisions," Economics Working Paper 2009-11.

VILMI, M. (2011): "The Effects of Endogenous Firm Exit on Business Cycle Dynamics and Optimal Fiscal Policy," Sveriges Riksbank Working Paper Series No. 250.

VII Figures and Tables

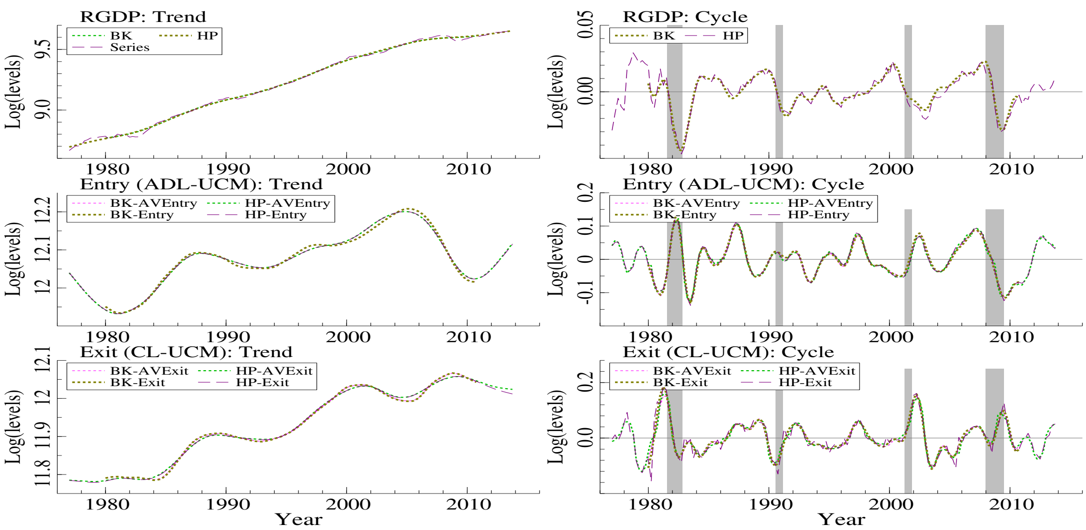
VIII Figures and Tables

Figure 1: The disaggregated series



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

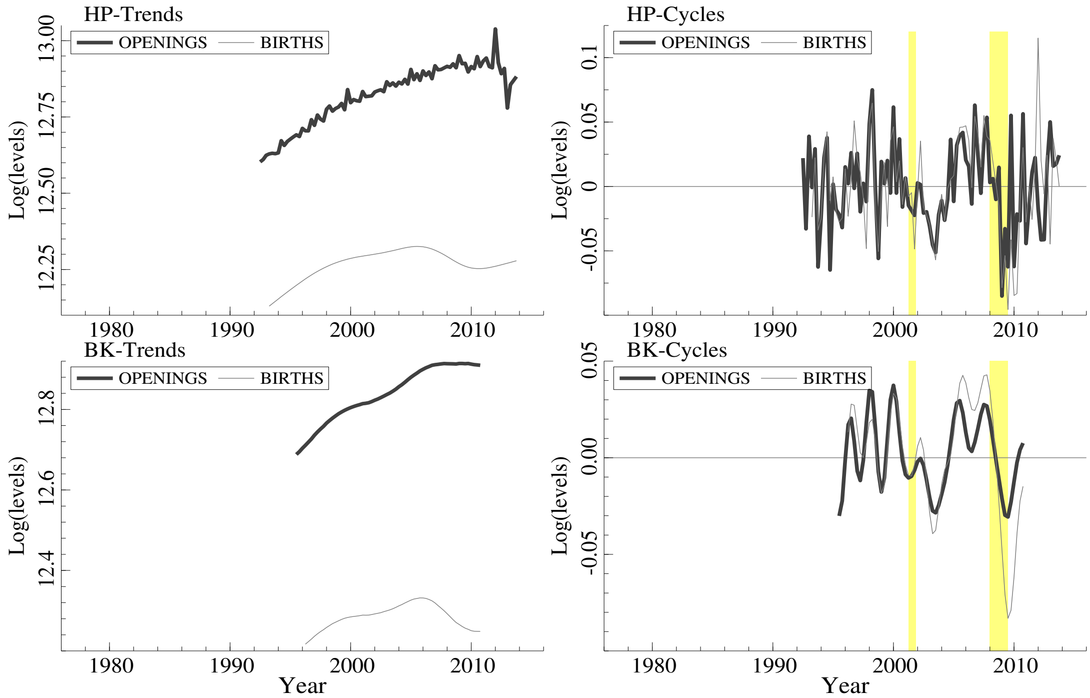
Figure 2: Business cycle analysis of disaggregated series



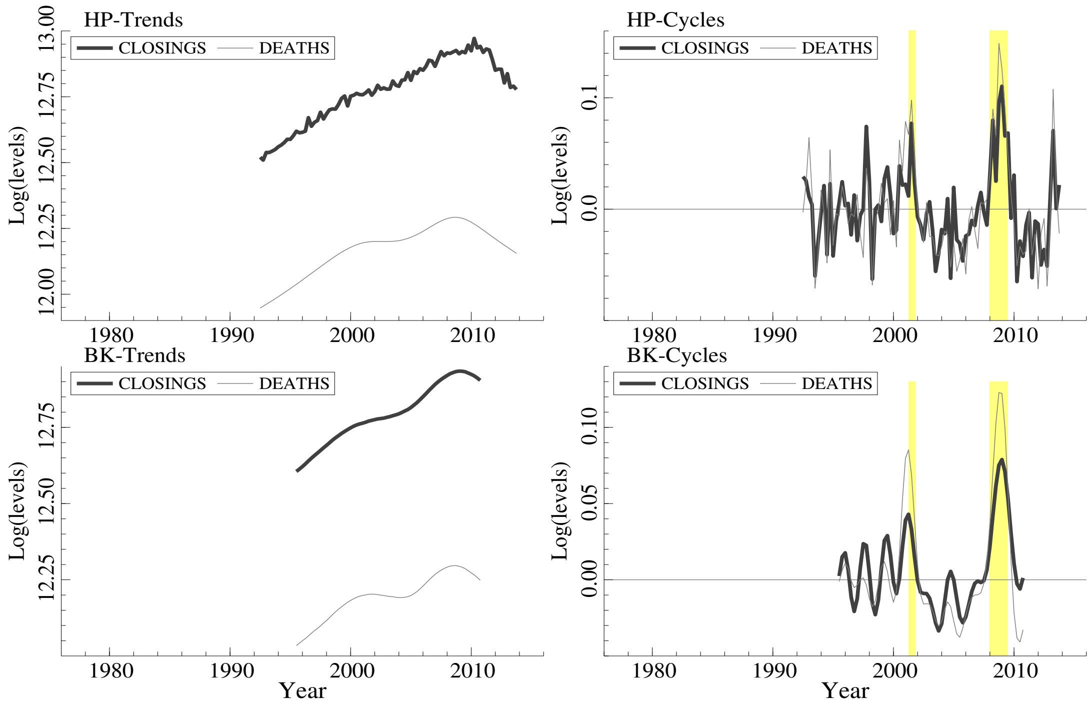
Notes: This figure shows the business cycle analysis of the RGDP (upper panel) and the ENTRY and EXIT series (central and lower panel, respectively) resulting from the application of univariate disaggregation discussed in Section III and combined UCM discussed in Section III. Namely, left (right) panels 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 3: Business cycle analysis of BLS quarterly series

(a) ENTRY

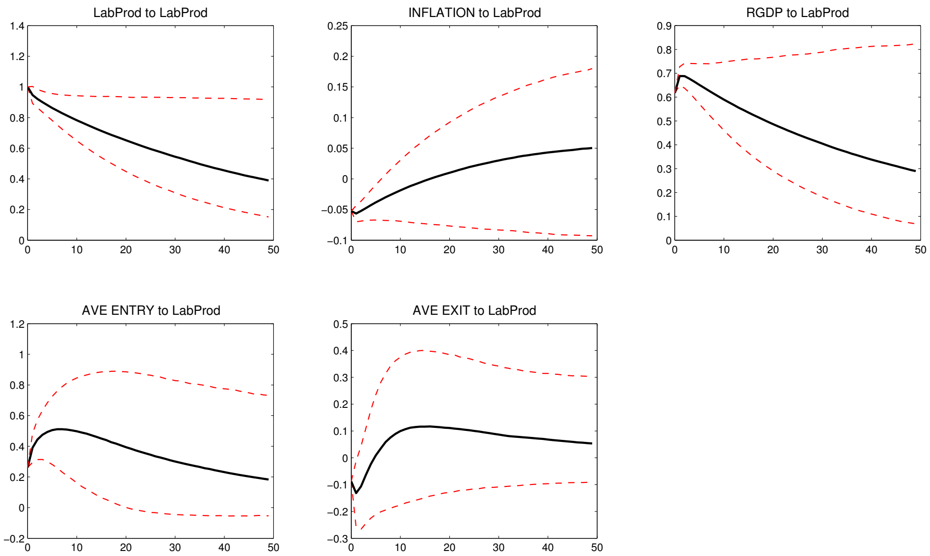


(b) EXIT



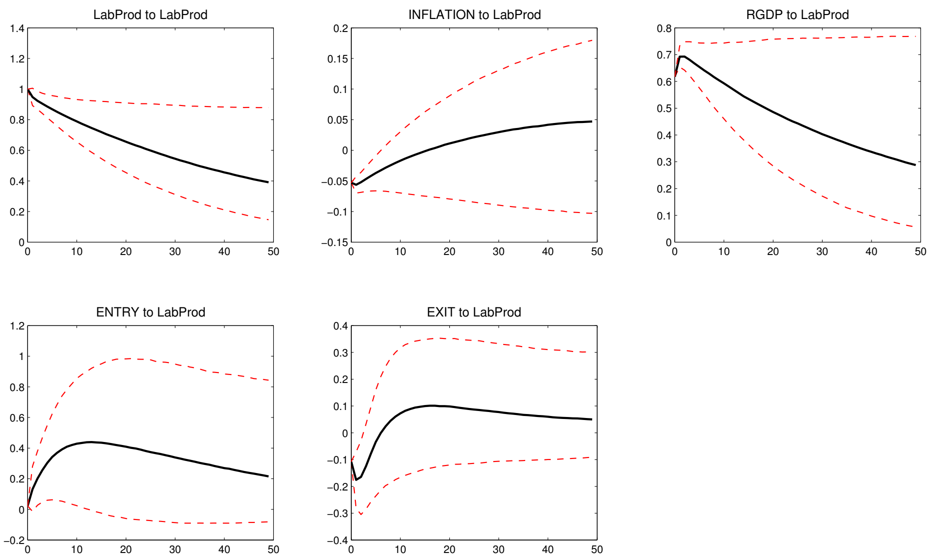
Notes: This figure shows the business cycle analysis of the BLS existing quarterly series of ENTRY – in panel (a) – and EXIT – in panel (b). Left (right) panels plot the trend (cycle) component extracted from the series in logarithms. Upper (lower) panels refer to HP (BK) filter. The cyclical components 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 4: IRFs from BVAR-M1 with Minnesota Priors.



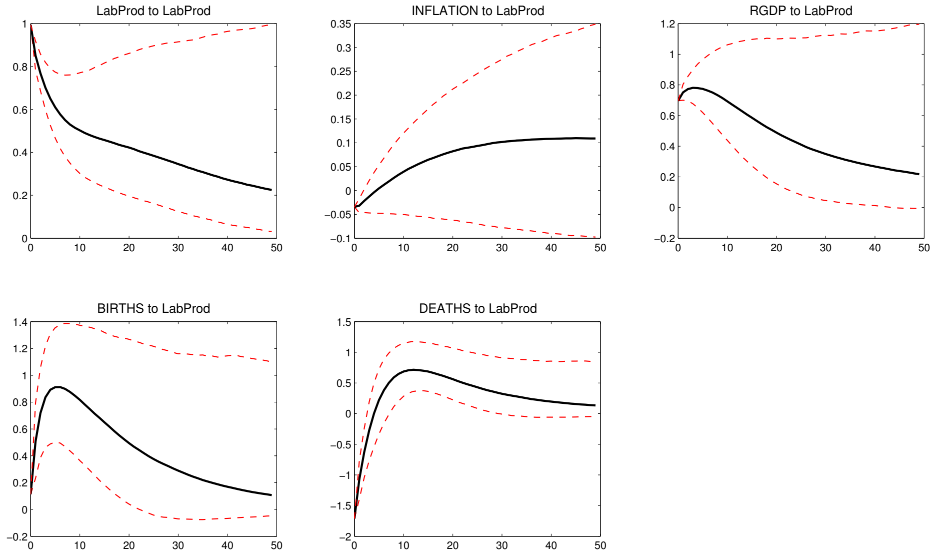
Notes: This figure shows the responses of the BVAR-M1 system to a labor productivity shock of one standard deviation. Sample: 1983Q1–2013Q4.

Figure 5: IRFs from BVAR-M2 with Minnesota Priors.



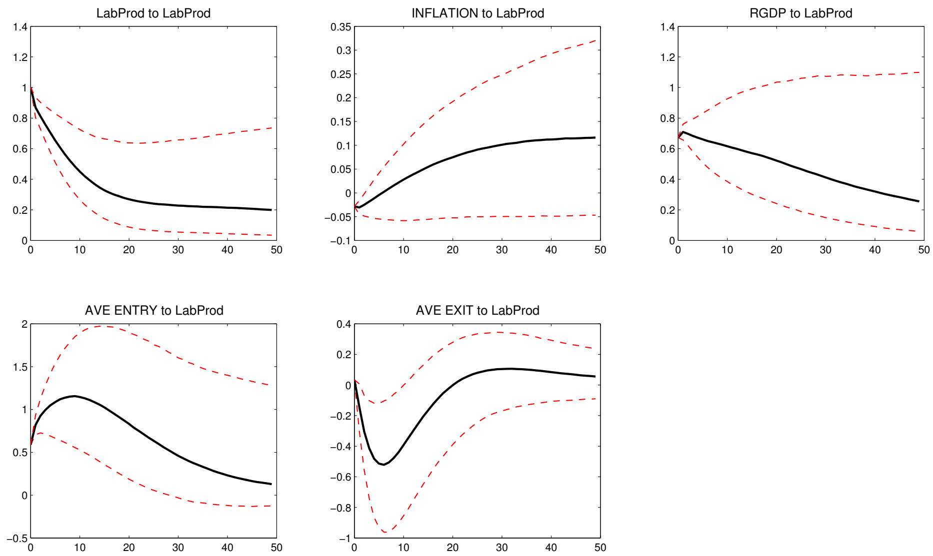
Notes: This figure shows the responses of the BVAR-M2 system to a labor productivity shock of one standard deviation. Sample: 1983Q1–2013Q4.

Figure 6: IRFs from BVAR-M3 with Minnesota Priors.



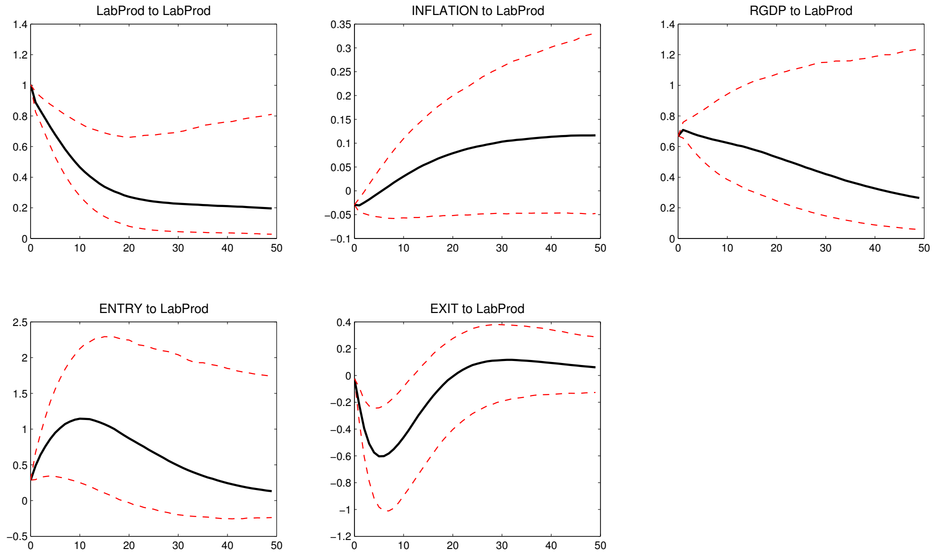
Notes: This figure shows the responses of the BVAR-M3 system to a labor productivity shock of one standard deviation. Sample: 1993Q3–2013Q4.

Figure 7: IRFs from BVAR-M4 with Minnesota Priors.



Notes: This figure shows the responses of the BVAR-M4 system to a labor productivity shock of one standard deviation. Sample: 1993Q3–2013Q4.

Figure 8: IRFs from BVAR-M5 with Minnesota Priors.



Notes: This figure shows the responses of the BVAR-M5 system to a labor productivity shock of one standard deviation. Sample: 1993Q3–2013Q4.

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	
Method	*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	

Notes: This table reports the correlation between the series resulting from disaggregation method accrding to (1) and the UCM according to equations (2)–(4) in Supplement with the specifications (3) and (4) and the existing quarterly data downloaded from BLS.

Table 2: BVAR Models description

Model	*Sample*	*Priors*	Y_t
M1	1983:Q1–2013:Q4	Minnesota	[LLabProd INF LRGDP LAVEntry LAVExit]'
M2			[LLabProd INF LRGDP LEntry LExit]'
M3	1993:Q2–2013:Q4	Minessota	[LLabProd INF DLRGDP BIRTHS DEATHS]'
M4			[LLabProd INF LRGDP LAVEntry LAVExit]'
M5			[LLabProd INF LRGDP LEntry LExit]'
M1b	1983:Q1–2013:Q4	Normal Diffuse	[LLabProd LCPI LRGDP LAVEntry LAVExit]'
M2b			[LLabProd LCPI LRGDP LEntry LExit]'
M3b	1993:Q2–2013:Q4	Normal Diffuse	[LLabProd LCPI LRGDP LBIRTHS LDEATHS]'
M4b			[LLabProd LCPI LRGDP LAVEntry LAVExit]'
M5b			[LLabProd LCPI LRGDP LEntry LExit]'

Notes: This table reports a list of definitions of vector $\mathbf{y_t}$ to be modeled by the BVAR system (8) for structural analysis. Namely, M1–M5 are BVAR estimated using Minnesota Priors. The labels M1b-M5b indicate that the BVAR models considering the Normal Diffuse Priors.

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	+

Notes: 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, $3 \times s$).