1. NET INDIRECT TAXES AND SECTORAL STRUCTURE OF ECONOMY[1]

Emilian Dobrescu[2]

Abstract

Usually, the sectoral structure of economy is measured as weights of the main branches: a) in the total gross value added or b) in the gross domestic product. Vector b differs from vector a by the sectoral net indirect taxes, as shown in the Input-Output Tables of Romania.

This issue has been explored using the Input-Output Tables of Romania for almost a quarter of a century. The primary information that resulted from the extended branch nomenclatures (from 90 to 105 positions in different years) has been aggregated into ten sectors. The series were methodologically homogenized according to the last Eurostat classification

The comparative analysis involved five structural coefficients derived from the Euclidean 1-norm distance, Bhattacharyya coefficient, Hellinger distance, Cosine similarity coefficient, and the so-called Jaccard index. Some computational problems of estimating - as autoregressive processes - the sectoral rates of the net indirect taxes are also examined.

Keywords: sectoral structure, net indirect taxes, VAR

JEL Classification: C53, C67, H2

I. Introduction

Usually, the sectoral structure of economy is characterized numerically by two vectors:

a. weights of the selected branches in the total gross value added (hereinafter denoted by wv_i) or b. the corresponding weights related, in this case, to the gross domestic product (wg_i).

The main differences between the mentioned vectors come from the sectoral distribution of the net indirect taxes. According to the national accounts definitions, the

---

[1] *The preliminary form has been presented as a working paper at "2014 IIOA Conference, Lisbon - Portugal (15-18th July 2014)".*

[2] *Centre for Macroeconomic Modelling, NIER, Romanian Academy. Email:emiliand@clicknet.ro*

net indirect taxes cumulate the value-added tax and excises, the customs duties, the public budget subsidies on product (with negative sign), and other similar taxes.

We shall consider the algebrical I-O relationships involved in this matter (I, j = 1, 2,…,n sectors):

$$wv_i = GVA_i/GVA \quad (1)$$

GVA_i – gross value added in the sector i, current prices

GVA – total gross value added, current prices

$$GVA_i = Q_i - \Sigma Q_j * a_{ji} \text{ for i fixed} \quad (2)$$

Q_i - output in sector i, current prices

a_{ji} – technical coefficients, current prices

$$GVA = \Sigma GVA_i \quad (3)$$

$$wg_i = GDP_i/GDP \quad (4)$$

GDP_i – gross domestic product in the sector i, current prices

GDP – total gross domestic product, current prices

$$GDP_i = GVA_i + NIT_i = GVA_i(1+rnit_i) \quad (5)$$

NIT_i – net indirect taxes in sector i, current prices

$rnit_i$ – rate of the net indirect taxes in sector I

$$GDP = \Sigma GDP_i \quad (6)$$

Normally, the rate of the net indirect taxes can be estimated not only sectorally ($rnit_i = NIT_i/GVA_i$), but also as an aggregate indicator (rnit=NIT/GVA).

This issue will be examined using, as a statistical application, the input-output tables of Romania for almost a quarter of a century (years 1989-2011), which offers a double advantage. On one hand, the series are methodologically homogenized for the entire period according to the last Eurostat classification (NCP 2013; NIS 2014). On the other hand, such an exercise is interesting because these series relate to a very dynamic structural process as that recorded by the Romanian economy during the transition from the centrally planned system to the functional market mechanisms.

To notice that the primary data resulted from the extended branch nomenclatures (90-105 positions in different years); these data have been aggregated into ten economic areas (Dobrescu 2013b).

This condensed structure shows the following (in brackets, the codes attached to the corresponding sectoral indicators):

- Agriculture, forestry, hunting, and fishing (1)
- Mining and quarrying (2)
- Production and distribution of electric and thermal power (3)
- Food, beverages, and tobacco (4)
- Textiles, leather, pulp and paper, and furniture (5)
- Machinery and equipment, transport means, and other metal products (6)
- Other manufacturing industries (7)
- Constructions (8)
- Transports, post, and telecommunications (9)
- Trade, business, and public service (10).
-

Appendix 1[3] "Statistical series" details, for 1989-2011 years, the available information concerning the gross value added, the net indirect taxes, and the gross domestic product in both aggregate and sectoral determination; the sectoral weights wvi and wgi, and the rates rniti are also determined.

Hereinafter, the paper is organized as follows.

Section II compares the structure of the Romanian economy based on GVA and GDP sectoral distributions. Thus, five structural measures (of what) (SC) are involved. They are derived from the following: the Euclidean 1-norm distance, the Bhattacharyya coefficient, the Hellinger distance, the Cosine similarity coefficient, and the so-called Jaccard index. These have been accommodated in such a way that the corresponding structural coefficients to be bounded by 0 (for all the forms of incongruity) and 1 (when the compared structures are identical). Our analysis reveals the important shifts occurred during transition in the sectoral configuration of the Romanian economy.

Section III examines some computational problems of estimating the sectoral rates of the net indirect taxes as a univariate autoregressive process. The attention is focused on the AR technique.

Several concluding statements close the paper.

II. Two Sectoral Structural Vectors

1. Our search attempts to approximate statistically the degree to which the sectoral structures of the Romanian economy in GVA and GDP determinations are - or are not - consonant.

Thus, five structural coefficients (SC) are estimated, according to formulas presented in Table 1 (in Dobrescu, 2011, pp.5-11, the necessary methodological considerations and documentary sources are presented *in extenso*). Normally, the symbols are adapted to our specific problem.

---

[3] *Appendix on http://rjef.ipe.ro*

Table 1

Computational formulas for the structural coefficient (SC)

Structural coefficient	Symbol	Formula
Euclidean 1-norm	SCE	$SCE = 1 - \frac{\sum \lvert wg_i - wv_i \rvert}{2}$
Bhattacharyya	SCB	$SCB = \sum \sqrt{wg_i \, wv_i}$
Hellinger	SCH	$SCH = 1 - \frac{\sqrt{\sum (\sqrt{wg_i} - \sqrt{wv_i})^2}}{\sqrt{2}}$
Cosine	SCC	$SCC = \frac{\sum w_i W_i}{\sqrt{\sum wg_i^2} \sqrt{\sum wv_i^2}}$
Jaccard	SCJ	$SCJ = \frac{\sum wg_i \, wv_i}{\sum wg_i^2 + \sum wv_i^2 - \sum wg_i \, wv_i}$

2. All these formulas have been applied to the series wv_i and wg_i of the Romanian I-O tables (Table 2).

Table 2

Structural Coefficients for the Period 1989-2011

Years	SCE	SCB	SCH	SCC	SCJ
1989	0.949214	0.997163	0.946741	0.992245	0.983774
1990	0.921931	0.993448	0.919056	0.981642	0.960321
1991	0.964262	0.998631	0.963002	0.997638	0.994697
1992	0.967219	0.996621	0.941869	0.996778	0.993576
1993	0.95726	0.998578	0.962292	0.996364	0.991635
1994	0.957574	0.998817	0.965602	0.996555	0.992319
1995	0.959807	0.998666	0.963476	0.997492	0.994091
1996	0.956984	0.998387	0.959838	0.997064	0.99341
1997	0.947413	0.997974	0.954986	0.99643	0.991168
1998	0.936326	0.9972	0.947082	0.994588	0.985047
1999	0.932077	0.996636	0.942002	0.994635	0.984195
2000	0.94314	0.997623	0.951249	0.996045	0.988803
2001	0.945742	0.997847	0.953602	0.996298	0.989627
2002	0.958111	0.998544	0.961847	0.997371	0.992224
2003	0.9481	0.997554	0.950548	0.996245	0.989028
2004	0.948163	0.997719	0.952239	0.996244	0.989336
2005	0.947448	0.997576	0.95077	0.996068	0.988235

Years	SCE	SCB	SCH	SCC	SCJ
2006	0.949103	0.997609	0.951101	0.996156	0.988433
2007	0.952309	0.997835	0.953468	0.996756	0.990016
2008	0.952975	0.997934	0.954543	0.996967	0.990537
2009	0.955043	0.997706	0.952101	0.996825	0.990629
2010	0.956205	0.998016	0.955455	0.996774	0.990976
2011	0.951135	0.99774	0.952456	0.995936	0.988105

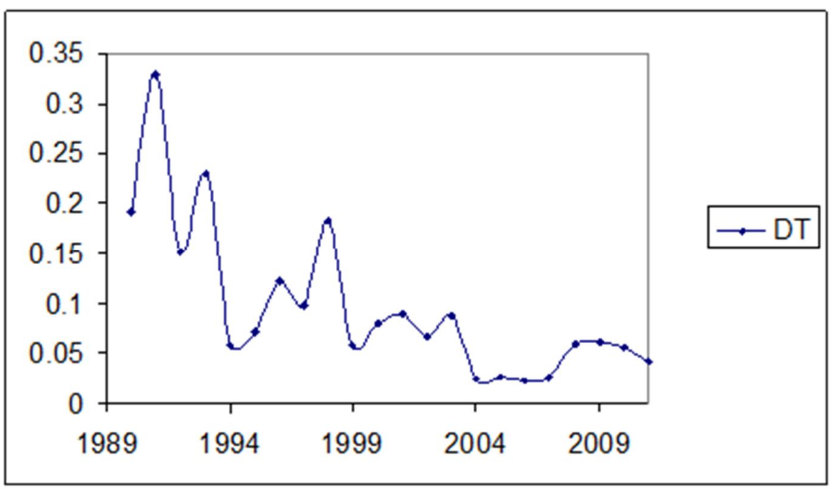
As expected, the sensitivity of these measures is not similar. For instance, the coefficient of variation (as a ratio of standard deviation to the sample mean) increases from 0.001093 for SCB and 0.003268 for SCC to 0.006922 for SCJ, reaching the highest levels in the case of SCH (0.010198) and SCE (0.010714). Despite these differences, all the measures indicate a noticeable dissimilarity between the examined two structural perspectives: gross value added and gross domestic product.

3. Obviously, such a discrepancy came from the modifications of the sectoral rates of the net indirect taxes ($rnit_i$). It must be mentioned that the intensity of these modifications was not uniform during the examined historical interval. It will be approximated by the annual changes coefficient (denoted by DT), computed as follows:

$$DT_t = [(1/n)*\Sigma_i(rnit_{it}-rnit_{i(t-1)})\hat{}2]\hat{}(1/2) \quad (7)$$

Figure 1

Annual Changes Coefficient



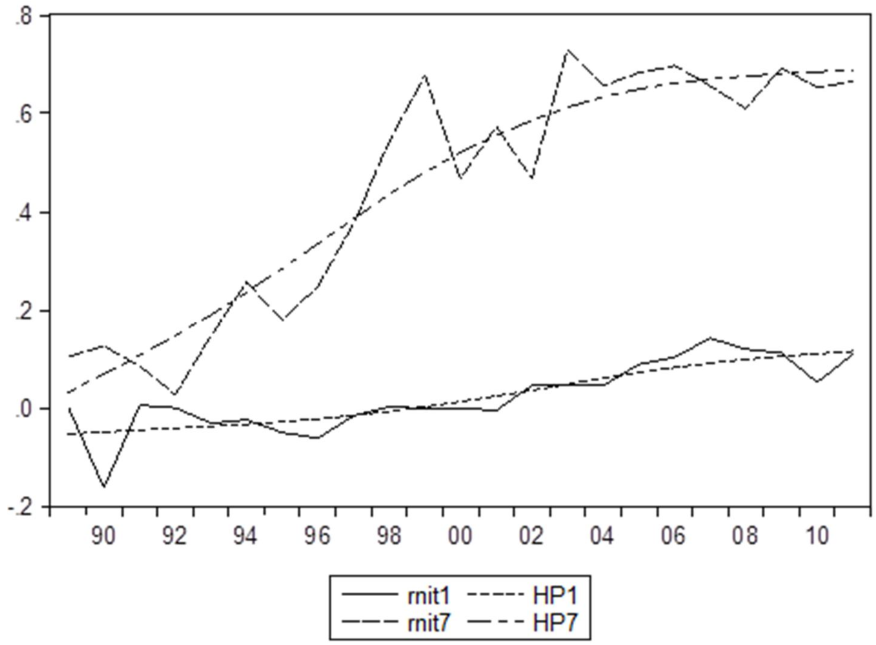
Generally, the first part of the interval is characterized by higher DT, which signifies more intensive changes in sectoral rates of the net indirect taxes. Such a temporal feature is understandable, taking into account that the initial phase of transition from the centrally planned to the market system inherently involves more extended institutional reforms, including the taxation.

4. Despite the annual volatility recorded by the sectoral NIT rates, several tendencies have been identified. To more clearly unfold them, the samples were filtered through the Hodrick-Prescott procedure (HP symbol).

4.1. An increasing trend characterizes the agriculture, forestry, hunting and fishing (sector 1), and other manufacturing industries (sector 7).

Figure 2

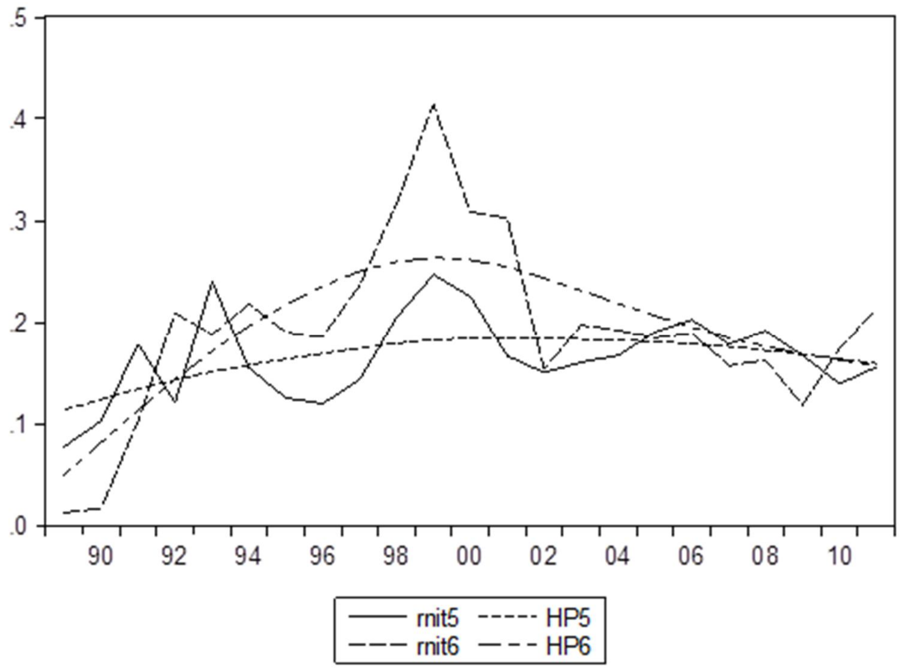
Ascendant Sectoral rNIT Rates



4.2. Relatively divergent evolutions appear in other fields. The Figure 3 covers the cases of initial expansion followed by ulterior diminishing of the net indirect taxes rates.

Figure 3

Ascendant-descendant Sectoral rNIT Rates

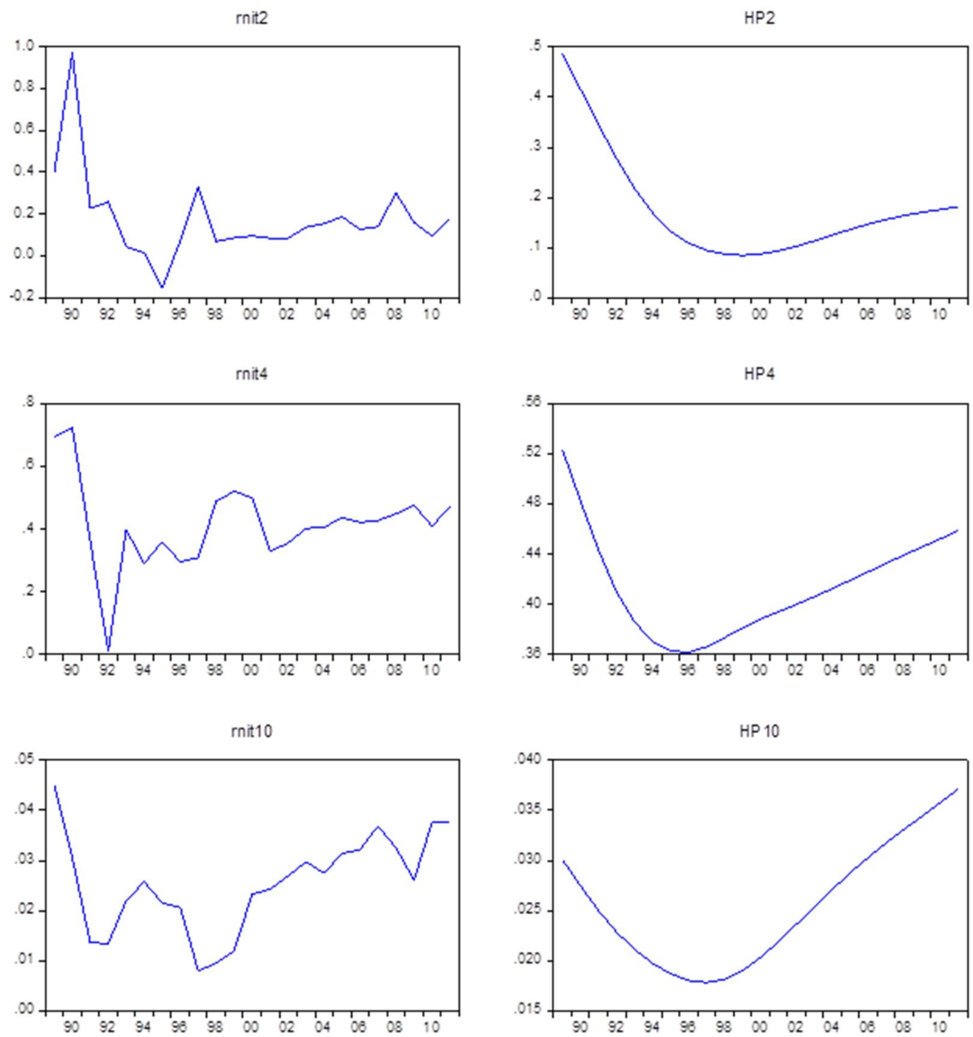


Such a pattern concerns the textiles, leather, pulp and paper, furniture (sector 5), machinery and equipment, transport means, and other metal products (sector 6).

An inverse picture can be seen (Figure 4), for instance, in mining and quarrying (sector 2); food, beverages, and tobacco (sector 4); and trade, business, and public service (sector 10).

Figure 4

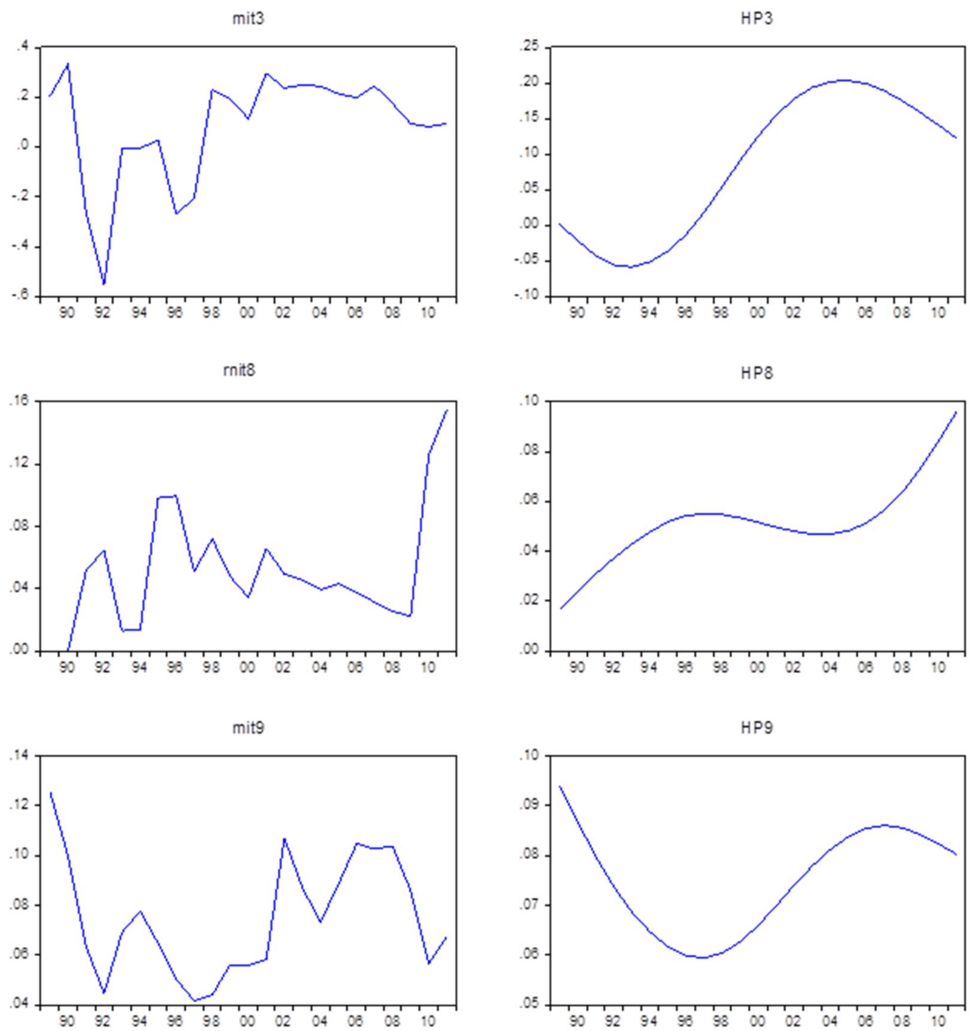
Descendant-ascendant Sectoral rNIT Rates



4.3. A quasi-cyclical dynamics characterizes the production and distribution of electric and thermal power (sector 3); construction (sector 8); and transports, post, and telecommunications (sector 9).

Figure 5

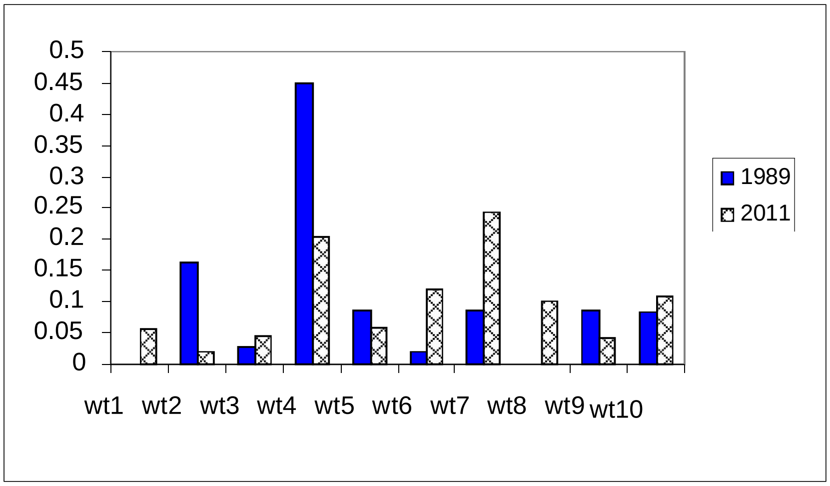
Oscillatory Sectoral rNIT Rates



4.4. As a consequence of these discrepancies, the sectoral structure of the net indirect taxes (wt_i) was significantly modified during the transition period (Figure 6).

Figure 6

Sectoral Structure of the Net Indirect Taxes Rates



Therefore, the relative contribution to the net indirect taxes of the Romanian economy has decreased in sectors 2, 4, and 9, this effect being compensated by its augmentation in sectors 1, 6, 7, and 8. The rest of economic activities (included in sectors 3, 5, and 10) underwent small modifications.

III. The Sectoral NIT Rates as Univariate AR Processes

The econometric estimation of the sectoral NIT rates could be interesting for both analytical and forecasting purposes. From the existing techniques, we shall pay attention in the present paper only to the univariate autoregressive algorithm.

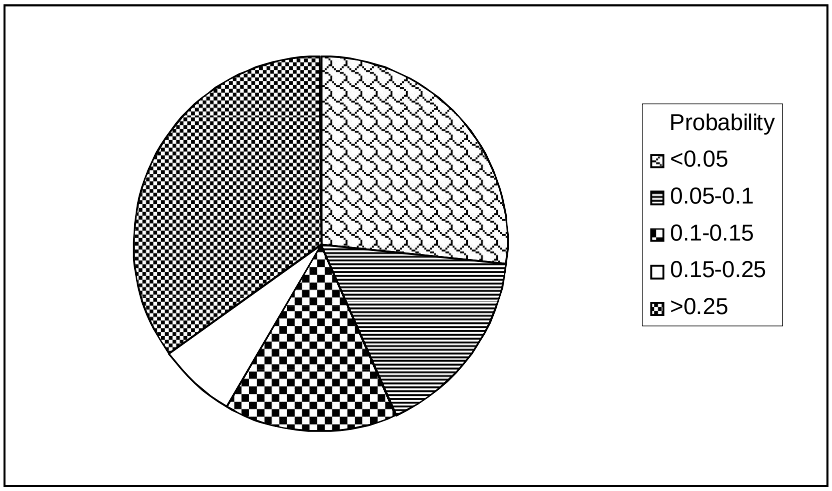
1. In the case of rniti series, the stationarity problem has not proven to be simple.

1.1. Three most usual unit root tests were involved: Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Elliott-Rothenberg-Stock (ERS). (Appendix 2 "Unit Root Test").

The first two have been computed in three variants: a) none, b) constant, and c) constant with linear trend, which means 60 statistics (10 statistical series, 2 tests, each of them in 3 exogenous variants). The probability to reject the null hypothesis (the series has a unit root) is distributed as in Figure 7.

Figure 7

Unit Root Tests



Therefore, in more than half of the cases, the probability to reject the null hypothesis exceeds 10%.

The Elliott-Rothenberg-Stock (ERS) test has been calculated for two variants: a) constant and b) constant with linear trend. The obtained results were also ambiguous.

1.2. Under such conditions, our computational strategy has been directed toward the stable AR, which involves the roots of the resulted characteristic polynomial that lie inside the unit circle (for details, see Lutkepohl 2007; Canova 2007; Uctum 2007; SCCN 2011; Rossi 2013; and Baum 2013). Such a solution can be considered relevant because “the same condition is necessary and sufficient for the stationarity of the stochastic process” (Nymoen 2013, p.41).

2. In AR specification, the lag length is essential. Consequently, during the last decades, this issue was of high interest for quantitative analysis and forecasting research. An illustrative list of the so-called optimal lag length selection is shown as follows:

- the simple graphical representation in Franz (1942) and Kunst (2007);
- the autocorrelation (AC) and the partial autocorrelation (PACF) of the given series in Dettling (2012) and Schwert (2013);
- Breusch-Godfrey LM or a Box-Ljung *Q* tests for residual autocorrelation in Parker (2014);
- the mean squared error (MSE) and the final prediction error (FPE) in Hafer and Sheehan (1987), Lutkepohl (2007), and Gupta and Miller (2012); the F-test in Hafer and Sheehan (1987);
- the Akaike Information Criterion (AIC) in Ozcicek and W. McMillin (1999), Dayton (2003), Canova (2007), Lutkepohl (2007), Gutierrez, *et al*. (2009), Gupta and Miller (2012), and Parker (2014);

- the Schwarz Information Criterion (SIC) in Hafer and Sheehan (1987), Ozcicek and W. McMillin (1999), Lutkepohl (2007), Canova (2007), Gutierrez, *et al*. (2009), Gupta and Miller (2012), and Parker (2014);
- the Hannan and Quinn Criterion (HQC) in Lutkepohl (2007), Canova (2007), Gutierrez, et al. (2009), Tarek (2012), and Gupta and Miller (2012);
- the Phillips' Posterior Information Criterion (PIC) in Phillips (1994) and Ozcicek and W.McMillin (1999).

Some variations of these procedures or more or less different approaches appear as the hypothesis and diagnostic tests in Kunst (2007); the sequential modified likelihood ratio (LR) in Gupta and Miller (2012); the Keating's modification of the AIC and SIC (KAIC and KSIC, respectively) in Ozcicek and W. McMillin (1999); and the Geweke's and Meese's (1981) Bayesian Estimation Criterion (BEC) in Hafer and Sheehan (1987). Dobrescu (2013a) used a composite structural inertiality index (SII) based on weighing aggregation of the AIC criterion, the length of post sample extrapolations up to convergence, and the coefficient of variation computed for this interval. Comprehensive considerations concerning the lag order selection can be found also in Burnham and Anderson (2002 and 2004), Dayton (2003), and Claeskens and Hjort (2008).

3. Preponderantly, these techniques take into account the similarities and differences between the fitted and primary data. Such a perspective is undoubtedly essential when a multivariate VAR is used to evaluate the separate contributions of the involved causal factors to the global dynamics of explained variable. A good approximation of available statistical data, especially of those concerning the pre-forecasting interval, would be also a more reliable tool for the short and medium-run prognoses.

However, the long-run behavior of a AR relationship can also be of interest not only from predictive reasons. In many cases, the statistical series could be "suspected" to contain some steady-state points or other time patterns. In such situations, the given VAR has to be submitted to a significant number of successive extrapolations to identify the possible stable dynamic features of the examined indicators.

This procedure has also been exercised on our example, for all the sectoral net indirect taxes rates (rniti) using nine lag lengths (from 2 to 10). Appendix 3 "AR econometrics" synthesizes the obtained results concerning the estimators, and R-squared.

Its analysis and, first of all, the repeated consecutive computations based on AR estimators reveal at least two important conclusions regarding the lag lengths.

3.1. As it was expected, some of these generate dynamically stable patterns, whereas others do not. In our application, we detected five types of long-run

- successive AR extrapolations (symbol in brackets):
- oscillatory asymptotic trend (OAT) behavior,
- oscillatory explosive trend (OET),
- smooth asymptotic trend (SAT),
- smooth explosive trend (SET), and
- erratic evolution (ERR).

From this point of view, the sectoral picture looks as follows (Table 3).

Table 3

Dynamic Patterns of the Successive Extrapolations Depending on the Adopted AR Length

Number of lags	2	3	4	5	6	7	8	9	10
rnit1	SAT	SAT	SAT	SAT	SET	OET	OAT	ERR	ERR
rnit2	SAT	OAT	OAT	OAT	OAT	OAT	OAT	OAT	OAT
rnit3	SAT	SAT	SAT	SAT	OAT	OAT	OAT	OAT	OAT
rnit4	OAT	OAT	SAT	SAT	OAT	OAT	OAT	OAT	ERR
rnit5	SAT	OAT	OAT	OAT	OAT	OAT	OAT	OAT	OAT
rnit6	SAT	OAT	SAT	OAT	OAT	OAT	OAT	OAT	OAT
rnit7	SAT	SAT	SAT	SAT	OAT	OAT	OAT	OAT	ERR
rnit8	OAT	OAT	OAT	OAT	OAT	OAT	OAT	ERR	ERR
rnit9	SAT	SAT	OAT	OAT	OAT	OAT	OAT	OAT	ERR
rnit10	SAT	SAT	SAT	SAT	OAT	OAT	OAT	ERR	ERR

Normally, the AR estimations characterized by oscillatory explosive trend (OET), smooth explosive trend (SET), and erratic evolution (ERR) cannot generate stable patterns of the long-run consecutive extrapolations. Eleven cases (underlined) are in such a situation.

3.2. The rest of AR relationships induce asymptotic (smooth or oscillatory) trends. Which one of them should be preferred? To answer this question, it would be useful to determine the stabilization interval of extrapolations, the number of successive AR computations until the asymptotical property of series is reached. If $extr_t$ and $extr_{(t-1)}$ represent two such extrapolations, the mentioned condition may be formulated as a rate $|extr_t/extr_{(t-1)}-1|<\varepsilon$, ε being considered as an acceptable maximal deviation from the asymptotical trend.

Regarding ε, the following two thresholds (0.01 and 0.001) have been admitted as significant in our analysis. Consequently, the following stabilization intervals were determined:

THR1 - the number of extrapolations, after which the consecutive rates become smaller than 0.01, and

THR2 - the number of extrapolations, after which the consecutive rates become smaller than 0.001.

The sectoral distribution of these parameters is represented in Table 4, which obviously contains information for cases having asymptotical properties (79 series).

Table 4

AR Stabilization Interval

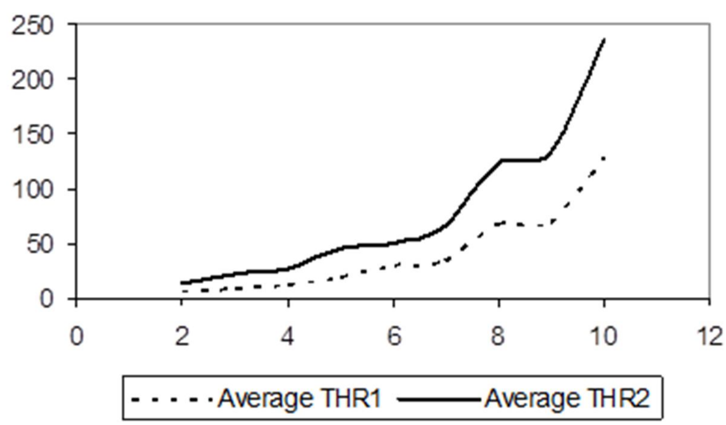
AR	Length	2	3	4	5	6	7	8	9	10
rnit1	THR1	8	17	20	43	***	***	170	***	***
	THR2	20	63	70	147	***	***	271	***	***
rnit2	THR1	4	7	8	27	25	16	15	19	71
	THR2	6	10	13	41	39	34	36	47	112
rnit3	THR1	6	6	11	12	28	55	23	25	31
	THR2	10	14	23	29	40	91	41	43	51

AR	Length	2	3	4	5	6	7	8	9	10
rnit4	THR1	5	8	4	4	14	9	11	37	***
	THR2	10	13	10	8	30	24	21	76	***
rnit5	THR1	3	6	23	37	43	33	19	22	27
	THR2	5	13	45	68	80	64	36	44	54
rnit6	THR1	2	7	8	17	22	24	92	81	381
	THR2	6	17	16	29	36	51	167	155	731
rnit7	THR1	3	2	7	3	8	11	14	19	***
	THR2	26	16	23	22	22	32	32	40	***
rnit8	THR1	21	27	26	40	98	53	68	***	***
	THR2	29	41	37	57	151	80	100	***	***
rnit9	THR1	5	6	7	5	16	80	221	276	***
	THR2	9	9	13	12	26	165	424	534	***
rnit10	THR1	8	8	3	7	12	21	50	***	***
	THR2	10	18	16	40	21	60	101	***	***

3.3. The data in Table 4 confirm the expected connection between the **AR** stabilization interval and the chosen **AR** length itself. To clearly reveal it, the Figure 8 represents the THR1 and THR2 (as averages for each **AR** length).

Figure 8

Averages of THR1 and THR2 for Each AR Length



The dependence of THR on the **AR** length is evident.

4. The methods mentioned under point 2 of this section seem unhelpful for the framework developed under point 3 regarding the choice of the AR length.

4.1. Summarizing, such an operation faces several problems.

4.1.1. Any econometric procedure aims to generate post-sample estimations that reproduce as truthfully as possible the properties of the reference statistics (mean, volatility). It is evident that the shorter the AR stabilization interval, the faster the extrapolations move away from the statistical characteristics of primary data. Therefore, the higher AR length would be preferable because they induce a longer stabilization interval.

The maximal lag length can be derived from the *degrees of freedom (df) specific to the given application, which* strongly depends on the available sample dimension (n) and the number of estimators involved in regression. *If the examined lag length is noted q, df is calculated simply as follows:*

$$df = n-2*q-2 \tag{8}$$

Normally, the maximum number of possibly usable AR lags (noted q_{max}) must correspond to the minimum number of degrees of freedom (df = 0), yielding the following:

$$n-2*qmax-2=0 \tag{9}$$

$$q_{max}=(n-2)/2=n/2-1 \tag{10}$$

Because the AR procedure operates only with integers, n will be decreased by 1 when it is odd.

4.1.2. On the other hand, with a relatively long extrapolating interval, the AR stability condition ceases to be of secondary importance as in the short- or medium-run applications. It becomes essential, attesting that the successive extrapolations based on the resulting estimators are convergent—they tend to a given level (steady state), or they oscillate within decreasing (or fixed) boundaries.

Besides, as already mentioned, such a requirement has also the role of the stationarity test.

4.2. As a result of observing both 4.1.1. and 4.1.2. points we obtain what could be called the longest stable vector of auto regression (LSVAR). How to define it practically? The simplest way is to verify the maximal AR length for stability condition. If this fails, the AR length is step-by-step compressed until the results cover the mentioned restriction.

5. Coming back to our application, the sectoral statistical series of the net indirect tax rates (rniti) contain 23 observations. This means that in all cases, the maximal AR length is 10. Appendix 3 "AR econometrics" presents the estimated VARs of rniti in all possible lag specifications (from 2 to 10). Table 5 reproduces the roots of the characteristic polynomials for the longest stable AR identified in the analyzed series.

Table 5

Roots of Characteristic Polynomials (Modulus) of the Longest Stable VARs in the rniti Series 1989-2011

Lag	rnit1	rnit2	rnit3	rnit4	Rnit5
(-1)	0.977252	0.935193	0.910685	0.940241	0. 10554
(-2)	0.977252	0.935193	0.910685	0.940241	0.910554

Lag	rnit1	rnit2	rnit3	rnit4	Rnit5
(-3)	0.942163	0.933229	0.90162	0.911385	0.900065
(-4)	0.900482	0.933229	0.844824	0.869292	0.900065
(-5)	0.900482	0.864963	0.844824	0.869292	0.865939
(-6)	0.541512	0.864963	0.774207	0.861118	0.865939
(-7)	0.541512	0.800475	0.774207	0.861118	0.73446
(-8)	0.428933	0.800475	0.754083	0.853231	0.73446
(-9)		0.797131	0.754083	0.853231	0.731575
(-10)		0.797131	0.387002		0.731575
Lag	rnit6	rnit7	rnit8	rnit9	rnit10
(-1)	0.992517	0.900374	0.933637	0.991009	0.951262
(-2)	0.992517	0.900374	0.933637	0.991009	0.951262
(-3)	0.970274	0.887024	0.926583	0.84953	0.947732
(-4)	0.970274	0.847574	0.926583	0.848435	0.947732
(-5)	0.94686	0.847574	0.86309	0.848435	0.911864
(-6)	0.94686	0.785971	0.86309	0.843348	0.911864
(-7)	0.945567	0.605041	0.816006	0.843348	0.904161
(-8)	0.945567	0.605041	0.816006	0.769596	0.904161
(-9)	0.914232	0.38414		0.769596	
(-10)	0.914232				

Note: No root lies outside the unit circle; AR satisfies the stability condition.

Therefore, LSVARs include the following:

10 lags in four series (rnit2, rnit3, rnit5, and rnit6),

9 lags in other three series (rnit4, rnit7, and rnit9), and

8 lags in the rest of three cases (rnit1, rnit8, and rnit10).

Normally, these results are corroborated with the data in Table 3 (to see the cells with highest number of lags before erratic evolutions).

6. The parameters of all computed LSVARs (constant and lag estimators) are synthesized in Table 6.

Table 6

AR Estimators for LSVARs

Constant and lag estimators	**rnit1**	**rnit2**	**rnit3**	**rnit4**	**rnit5**
c	0.013471	0.170334	0.096398	0.20311	0.330669
(-1)	0.532626	0.135147	0.334062	0.281894	0.536233
(-2)	0.435684	-0.1547	0.106355	-0.14719	-0.77999
(-3)	-0.41788	0.541103	0.067918	0.248268	0.097871
(-4)	0.517527	-0.21198	0.058012	-0.3668	-0.49716

Constant and lag estimators	rnit1	rnit2	rnit3	rnit4	rnit5
(-5)	0.042292	-0.08745	-0.14001	0.246694	0.076921
(-6)	-0.16652	-0.17985	0.15539	0.045718	-0.1078
(-7)	-0.23783	-0.01893	-0.2343	0.055744	-0.1211
(-8)	-0.09177	-0.04879	0.071157	-0.12608	-0.01058
(-9)		-0.01031	-0.10708	0.328678	0.095292
(-10)		-0.23202	-0.0704		-0.14541
Constant and lag estimators	**rnit6**	**rnit7**	**rnit8**	**rnit9**	**Rnit10**
C	0.595599	0.658895	0.108653	0.07954	0.010791
(-1)	-0.22253	-0.29317	0.689524	0.554027	0.647532
(-2)	0.149745	0.206885	-1.07769	-0.35067	0.159752
(-3)	-0.28242	-0.2953	0.592245	0.409597	-0.11882
(-4)	-0.22835	-0.13773	-0.956	-0.08414	0.12397
(-5)	0.161875	0.401132	0.522418	-0.10798	-0.00492
(-6)	-0.31617	0.134139	-0.64766	0.131911	-0.04889
(-7)	-0.15992	-0.06076	0.163861	-0.40158	0.41142
(-8)	-0.04394	0.135079	-0.37121	0.004727	-0.55249
(-9)	-0.1602	-0.0571		-0.25299	
(-10)	-0.62135				

7. The residuals (symbol $RESrnit_i$) of the estimated LSVARs were submitted to the normality and autocorrelation tests. In all the cases, the lag length is long enough (8-10 terms) comparatively to the given statistical series, which obviously reduces the power of the tests. Nevertheless, they were considered useful, at least as auxiliary information.

7.1. The Jarque-Bera procedure has been performed as a normality test (Table 7).

Table 7

Jarque-Bera Test for LSVARs' Residuals

	RESrnit1	RESrnit2	RESrnit3	RESrnit4	RESrnit5
Skewness	0.1673	1.705819	-0.262339	1.31698	0.498615
Kurtosis	3.090865	5.993603	2.691461	5.02537	2.450946
Jarque-Bera	0.075133	11.15884	0.200678	6.43989	0.70196
Probability	0.96313	0.003775	0.904531	0.03996	0.703998
	RESrnit6	RESrnit7	RESrnit8	RESrnit9	RESrnit10
Skewness	-0.1845	0.166849	0.52213	1.21861	0.14709
Kurtosis	1.732361	1.842295	3.633237	4.55382	2.992022
Jarque-Bera	0.944165	0.846787	0.932167	4.87337	0.054128
Probability	0.623702	0.654821	0.627455	0.08745	0.973299

Therefore, the normality hypothesis cannot be accepted only in two cases: RESrnit2 and RESrnit4; for RESrnit9, this probability is also high. To be sure, these three series were checked using other normality tests, namely, sfrancia (Shapiro-Francia) and Shapiro-Wilk scores. They confirmed again the Jarque-Bera results.

Anyhow, it is a fact that for the great majority of LSVARs' series (seven from ten), the normality distribution of residuals cannot be rejected, which justifies their utilization.

7.2. Concerning the autocorrelation of LSVARs' residuals, the LM test has been applied (Table 8).

Table 8

LM Test for LSVARs' Residual Serial Correlation

Lags	RESrnit1		RESrnit2		RESrnit3		RESrnit4		RESrnit5
	LM-Stat	Prob	LM-Stat	Prob	LM-Stat	Prob	LM-Stat	Prob	LM-Stat
1	0.09738	0.755	0.11115	0.7388	0.14279	0.7055	4.96154	0.0259	0.77014
2	0.00108	0.9737	0.53249	0.4656	0.11	0.7401	0.0489	0.825	0.15563
3	0.38665	0.5341	2.51846	0.1125	1.23835	0.2658	0.32952	0.5659	0.00174
4	1.44654	0.2291	0.00466	0.9456	0.03303	0.8558	1.08153	0.2984	0.3004
5	1.72479	0.1891	0.2597	0.6103	0.13783	0.7105	0.83583	0.3606	0.98422
6	0.50003	0.4795	0.96492	0.326	0.04441	0.8331	0.17592	0.6749	1.45182
7	0.03829	0.8449	0.97567	0.3233	0.03529	0.851	0.63693	0.4248	0.66552
8	1.60206	0.2056	0.01508	0.9023	0.02872	0.8654	1.12843	0.2881	0.62543
9			2.952	0.0858	0.02511	0.8741	1.16172	0.2811	2.02848
10			1.51983	0.2176	0.07313	0.7868			2.79634
Lags	RESrnit6		RESrnit7		RESrnit8		RESrnit9		RESrnit10
	LM-Stat	Prob	LM-Stat	Prob	LM-Stat	Prob	LM-Stat	Prob	LM-Stat
1	0.37538	0.5401	0.94403	0.3312	0.00178	0.9664	2.48685	0.1148	0.03502
2	1.0392	0.308	0.00175	0.9666	0.31933	0.572	2.70695	0.0999	0.06345
3	0.24081	0.6236	0.539	0.4628	0.39769	0.5283	4.52338	0.0334	1.50684
4	0.00816	0.928	0.02847	0.866	5.24053	0.0221	0.24571	0.6201	0.53516
5	0.06322	0.8015	3.96153	0.0466	0.01264	0.9105	3.76938	0.0522	4.18136
6	0.15857	0.6905	0.47153	0.4923	2.26869	0.132	0.02252	0.8807	0.55029
7	0.02004	0.8874	0.29164	0.5892	1.7779	0.1824	0.02796	0.8672	1.52E-05
8	0.0346	0.8524	0.02674	0.8701	2.75839	0.0967	1.75287	0.1855	2.25024
9	1.02372	0.3116	0.1653	0.6843			0.13989	0.7084	
10	0.36338	0.5466							

Generally, the serial autocorrelation of residuals does not seem to be a problem in the case of LSVARs series. This corroborates Parker's (2014) remark: "adding lags…to the right-hand side of a distributed-lag regression usually lessens the degree of autocorrelation in the error term (p.54)." Our series really are long enough comparatively to the available sample.

8. We shall illustrate the acceptability of LSVARs' estimations from two final applicative criteria.

The first concerns their ability to approximate adequately the statistical properties of the involved sample.

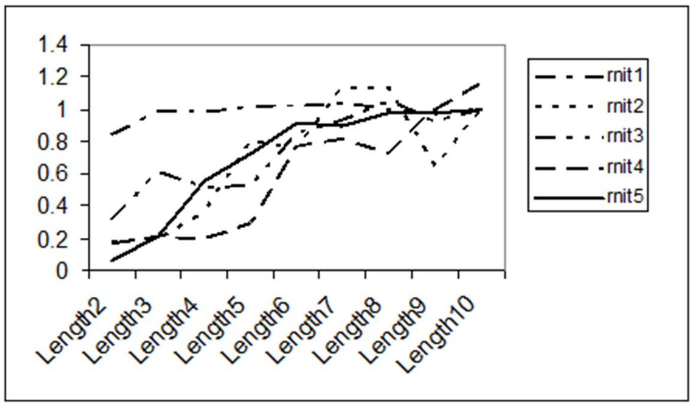
The second refers to the pattern induced by the long-run successive extrapolations, in connection with the stability of the AR condition.

8.1. Relating to the first, we shall focus on R-squared. As already outlined, in our application, longer possible AR models are preferred. Consequently, the adjusted R-squared seems to be irrelevant. (see Appendix 3 “AR econometrics”) To easily interpret the data, R-squared of all AR lengths (from 2 to 10) were scaled against the R-squared of LSVARs (=1).

8.1.1. The Figure 9 presents the coefficients of determination for sectors 1-5.

Figure 9

Coefficients of Determination for Sectors 1-5

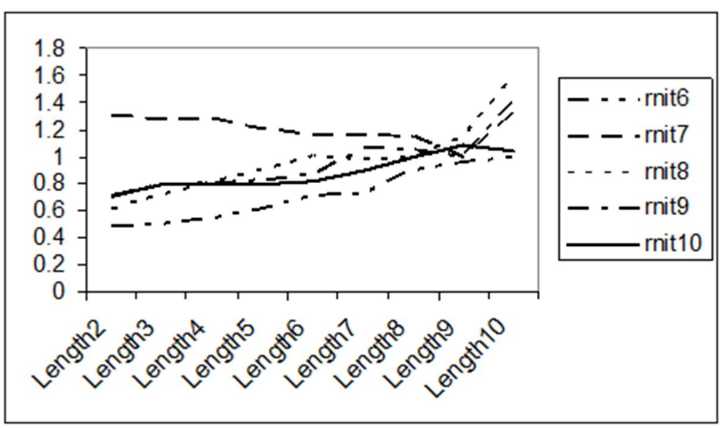


Therefore, there were R-squared higher than those of the LSVARs in only 8 cases.

8.1.2. We shall proceed similarly for the other five sectoral net indirect tax rates (rnit6-rnit10).

Figure 10

Coefficients of Determination for Sectors 6-10



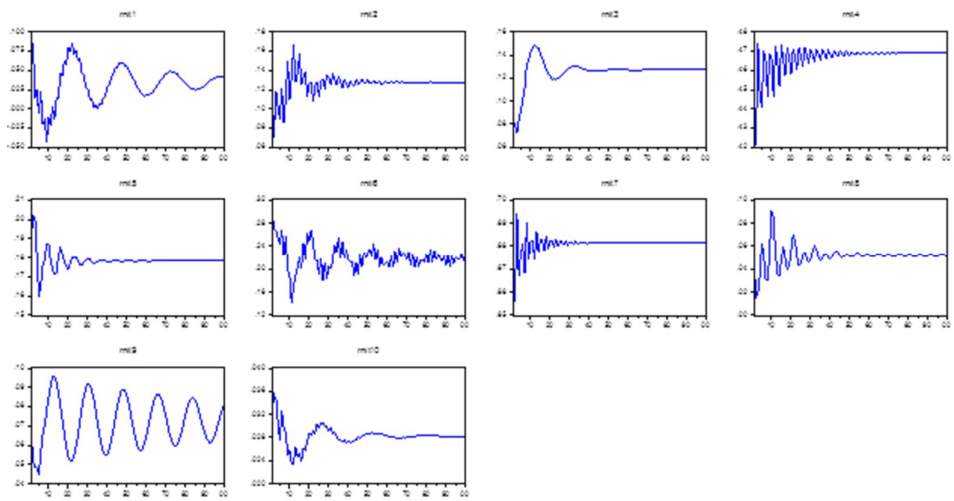
In the second group, the number of series with greater R-squared (than LSVARs) is larger, remaining around one third.

Synthetically, from a total of 90 estimations (ten series regressed for nine lag lengths), only approximately one quarter have R-squared higher than LSVARs. In our opinion, such a proportion indicates an acceptable degree of LSVARs to approximate the statistical properties of the involved sample.

8.2. The post-sample simulations have been computed for different intervals. The following graphs retain dynamics that resulted from 100 successive extrapolations for all ten series $rnit_i$.

Figure 11

Results from 100 Successive Extrapolations for All Series of rniti



The asymptotical property of all estimated LSVARs is also graphically illustrated.

IV. Final Remarks

1. The paper examined the sectoral structure of economy measured as weights of the main branches in the total gross value added and in the gross domestic product. As a numerical example, the input-output tables for Romania (annual data for the period 1989-2011, aggregated into ten sectors) were used.

The comparative analysis involved five structural coefficients (SCs) derived from the Euclidean 1-norm distance, Bhattacharyya coefficient, Hellinger distance, Cosine similarity coefficient, and the so-called Jaccard index. All of them indicated some dissimilarities between the two mentioned structural perspectives, induced by the sectoral distribution of the net indirect taxes rates ($rnit_i$, computed as ratios of the sectoral net indirect taxes to the corresponding gross value added).

2. This distribution registered important shifts during the transition from centrally planned economy to market system, especially in the first part of the interval, when initial breaking institutional reforms were promoted. Despite the volatility registered by the sectoral $rnit_i$, several tendencies have been however identified. An increasing trend characterizes sectors 1 (agriculture, forestry, hunting, and fishing) and 7 (other manufacturing industries). In sectors 5 (textiles, leather, pulp and paper, and furniture) and 6 (machinery and equipment, transport means, and other metal products), the net indirect tax rates initially expanded, followed by a downward trend. An inverse picture characterized sectors 2 (mining and quarrying), 4 (food, beverages, and tobacco), and 10 (trade, business, and public service). A quasi-cyclical pattern could be seen in the case of sectors 3 (production and distribution of electric and thermal power), 8 (constructions), and 9 (transports, post, and telecommunications).

3. The econometric estimation of the sectoral $rnit_i$ becomes necessary in any analysis or prognosis of the sectoral structure of economy. Obviously, such an objective could be reached in different ways. The present paper focused only on the univariate autoregressive algorithm; further studies will try other approaches.

For available series, the stationarity problem has not proved to be simple. Three of the most usual unit root tests were applied: Augmented Dickey-Fuller, Phillips-Perron, and Elliott-Rothenberg-Stock. However, the results were ambiguous. Consequently, the computational strategy has focused on the stable VAR, which shows that the roots of the resulted characteristic polynomial must lie inside the unit circle.

4. In choosing the AR lag length, the paper has insisted on both targetable goals in such estimations.

Undoubtedly, the coefficient of determination of regression cannot be ignored. A good approximation of available statistical data, especially of those that refer to the preforecasting interval, represents a reliable tool for the short- and medium-run prognoses.

However, the long-run behavior of a AR relationship can also be of interest not only because of predictive reasons. In many cases, the statistical series could be "suspected" to contain some steady-state points or other time patterns. In such situations, the given AR has to be submitted to a significant number of successive extrapolations to identify the possible stable dynamic features of the examined indicators.

5. The paper introduces the so-called "the longest stable vector of auto regression (LSVAR)," based on two premises:

It starts from the maximal AR lag length (q_{max}), which corresponds to the minimum number of degrees of freedom ***specific to the respective application, dependent in its turn on*** the available sample dimension (n data). Thus, $q_{max}=(n/2-1)$ is derived; because the VARs operate only with integers, n will be diminished by 1 when it is odd.

On the other hand, the AR stability condition is also essential. It proves that the successive extrapolations based on the resulting estimators are convergent, tending to a given level (steady state point) or to fixed or decreasing boundaries of oscillations.

The practical procedure begins by checking for stability condition of the maximal AR length. If such a test fails, the AR length is step-by-step compressed until the mentioned restriction is observed. The examined paper statistical series ($rnit_i$) contain 23 observations, which means that in all cases, the maximal AR length equals to 10 lags. Only in four cases, such a length corresponds also to the stability condition: rnit2, rnit3, rnit5, and rnit6. LSVARs include 9 lags in other three series (rnit4, rnit7, and rnit9) and 8 lags in the rest of the three cases (rnit1, rnit8, and rnit10).

6. The obtained LSVARs were analyzed, taking into account several econometric and applicative criteria.

The residuals of regression were submitted to the normality and autocorrelation tests. Obviously, their power is lessened by the circumstance that the lag length in all cases is sufficiently long (8-10 terms) comparatively to the statistical sample. Nevertheless, such an exercise has been considered useful at least as auxiliary information. For the great majority of LSVARs' series (seven from ten), the normality distribution of residuals cannot be rejected, which justifies their utilization. The LM test also shows that the autocorrelation of residuals does not seem to be a disquieting problem in the case of these series.

The R-squared coefficient was also admitted as relevant. Synthetically, from a total number of 90 estimations (ten series regressed for nine lag lengths), only approximately one quarter have R-squared higher than LSVARs. Such a proportion indicates, in our opinion, an acceptable degree of LSVARs to approximate the statistical properties of the involved sample.

The post-sample simulations (computed for different intervals of successive extrapolations) confirmed the asymptotical properties of all LSVARs estimated in the present paper.

7. LSAR steady-state level of $rnit_i$ (noted $srnit_i$) has been calculated as a mean of 100 estimations post-THR2, that is, after which, the relative change of two consecutive extrapolations became lower than 0.001.

Table 9

Steady-state Level of rniti Resulting from LSVAR

srnit1	srnit2	srnit3	srnit4	srnit5
0.034915	0.12743	0.127039	0.468994	0.17819
srnit6	srnit7	srnit8	srnit9	srnit10
0.218709	0.681511	0.052124	0.072495	0.028218

The net indirect tax rates, as steady-state levels, are positive in all the sectors.

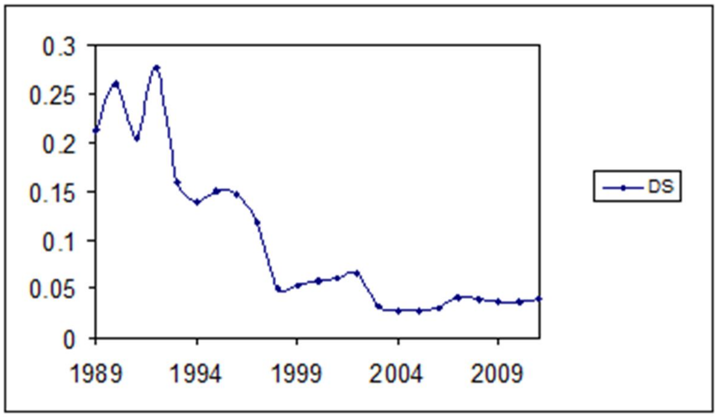
It would be interesting to compare the actual evolution of $rnit_i$ with their steady-state levels. The DS will be used as a synthetic indicator with such a goal:

$$DS=(\Sigma wv_i*(rnit_i-srnit_i)^\wedge 2)^\wedge 0.5 \qquad (11)$$

We remind that wv_i represents the sectoral weights of the gross value added. The Figure 12 describes the trajectory of this indicator during the period 1989-2011.

Figure 12

Deviation of Actual Evolution of rniti with Their Steady-State Levels



Further research has to check if such a picture is confirmed by other econometric techniques than LSVAR.

Acknowledgments

The author thanks I. Ghizdeanu, V. Gaftea, and T. Bria for the valuable assistance in building the data base used in the present paper.

References

Baum, C.F., VAR, SVAR and VECM models. EC 823: *Applied Econometrics*, Boston College, Spring 2013. Available at: <http://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn10.slides.pdf> [Accessed February 28, 2014].

Burnham, K.P. and Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information - Theoretical Approach*. Second Edition, New York: Springer-Verlag.

Burnham, K.P. and Anderson, D.R., 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), pp. 261-304.

Canova, F., 2007.Methods for Applied Macroeconomic Research.Princeton University Press. Available at: <http://www.eui.eu/Personal/Canova/Articles/ch4.pdf> [Accessed December 23, 2013].

Claeskens, G. and Hjort, N.L., 2008. Model Selection and Model Averaging.*Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.

Dayton, C.M., 2014. Model Comparisons Using Information Measures.*Journal of Modern Applied Statistical Methods*, 2(2), pp. 281-292. Available at: <http://education.umd.edu/EDMS/fac/Dayton/PCIC_JMASM.pdf>[Accessed January 26].

Dettling, M., 2012.Applied Time Series Analysis. Institute for Data Analysis and Process Design, Zurich University of Applied Sciences. <http://stat.ethz.ch/education/semesters/ss2012/atsa/ATSA-FS12-Week06.pdf> [Accessed December 1, 2013].

Dobrescu, E., 2011. Sectoral Structure and Economic Growth.*Romanian Journal of Economic Forecasting*, 14(3), pp. 5-36.

Dobrescu, E., 2013a. Modelling the sectoral structure of the final output.*Romanian Journal of Economic Forecasting*, 16(3), pp. 59-89.

Dobrescu, E., 2013b. Updating the Romanian Economic Macromodel.*Romanian Journal of Economic Forecasting*, 16(4), pp. 5-31.

Franz, L., 1942. Distributed Lags.*Econometrica*, 10(2), pp. 113-128, The Econometric Society Stable URL: <http://www.jstor.org/stable/1905784> [Accessed September 19, 2013].

Gupta, R. and Miller, S.M., 2012. <Ripple effects> and forecasting home prices in Los Angeles, Las Vegas, and Phoenix. *The Annals of Regional Science*, 48(3), pp. 763-782. Available at: <http://www.dspace.up.ac.za/bitstream/handle/2263/19836/Gupta_Ripple(2012).pdf?sequence=1> [Accessed January 23, 2014].

Gutierrez, C.E.C. Souza, R.C. and Guillén, O.T.C., 2014. Selection of optimal lag length in cointegrated VAR models with weak form of common cyclical features. *MPRA Paper* 22550.

Hafer, R.W. and Sheehan, R.G., 1987. On the Sensitivity of VAR Forecasts to Alternative Lag Structures.Federal Reserve Bank of St. Louis - Research Division, *Working Papers Series*, Working Paper 1987-004A. Available at: <http://research.stlouisfed.org/wp/1987/1987-004.pdf>[Accessed January 22].

Kunst, R.M., 2007. Vector auto regressions. Available at: <http://homepage.univie.ac.at/robert.kunst/var.pdf> [Accessed December23 2013].

Lutkepohl, H., 2007. *New Introduction to Multiple Time Series Analysis*.Springer.

NCP - National Commission for Prognosis Romania, 2013. Project “Strengthening the institutional capacity in evaluating and formulating of macroeconomic policies for economic convergence with EU within the National Commission for Prognosis”, SMIS code 27153. <http://www.cnp.ro/inovatie/docs/conferinta-de-inchidere-20-03-2013/04_Prezentare%20Academician%20Dobrescu-final%20romana.pdf>.

NIS – National Institute for Statistics Romania, 2014. Input-Output Tables.TEMPO- <https://statistici.insse.ro/shop/index.jsp?page=tempo2&lang=ro&context=35>.

Nymoen, R., 2013. GLS.Concepts of dynamic modeling. ECON 4160, Spring term 2013, Lecture 3, University of Oslo. Available at: <http://www.uio.no/studier/emner/sv/oekonomi/ECON4160/h13/undervisningsmateriale/lect3h13.pdf> [AccessedJanuary 28, 2014].

Ozcicek, O. and McMillin, W.D., 1999. Lag Length Selection in Vector Autoregressive Models: Symmetric and Asymmetric Lags. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.4267&rep=rep1&type=pdf> [Accessed December 23, 2013].

Parker, J.A., 2014.*Economics 312*: Theory and Practice of Econometrics. Chapter 3, Reed College Portland, Oregon. Available at: <http://academic.reed.edu/economics/parker/s13/312/tschapters/S13_Ch_3.pdf> [Accessed January 23, 2014].

Phillips, P.C.B., 1994. Model Determination and Macroeconomic Activity.Cowles Foundation for Research in Economics, *Discussion Paper* No. 1083. Available at: <http://cowles.econ.yale.edu/P/cd/d10b/d1083.pdf> [Accessed January 23, 2014].

Rossi, E., 2013. Vector Auto Regression Model.University of Pavia. Available at: <http://economia.unipv.it/pagp/pagine_personali/erossi/Econometria_Finanziaria_2013/Rossi_VAR_Ec_Fin_2013.pdf> [Accessed December 23, 2013].

SCCN - Swartz Center for Computational Neuroscience, 2011. *SIFT Online Handbook and User Manual*, Chapter 3.1. Stationarity and Stability.University of California San Diego. Available at: <http://sccn.ucsd.edu/wiki/Chapter_3.1._Stationarity_and_Stability> [Accessed February 28, 2014].

Schwert, G.W., 2013. Time Series Analysis: ARIMA Models. APS 425 - *Advanced Managerial Data Analysis*. Available at: <http://schwert.ssb.rochester.edu/a425/a425_arma.pdf> [Accessed November 30, 2013].

Tarek, J., 2012. Strongly consistent order selection procedures for long auto regression fitting.*Review of Business Research*, 12(3), pp. 44-69

Uctum, M., 2007. Economics 82400: Applied Macroeconometrics - Stability Analysis for VAR systems. CUNY Graduate Center. Available at: <http://userhome.brooklyn.cuny.edu/economics/muctum/EconometricsG/Stability%20Analysis%20for%20VAR%20systems.doc> [Accessed December 23, 2013]