Predicting Gross Domestic Product Components through Tsallis Entropy Econometrics

S. Bwanakare[a], M. Cierpiał-Wolan[b] and A. Mantaj[a]

[a]University of Information Technology and Management (WSIZ), Rzeszów, Poland
[b]Statistics Office in Rzeszów, University of Rzeszów, Poland

This article proposes the Tsallis non-extensive entropy econometric approach to forecast components of the country gross domestic product based on the knowledge of time series macroeconomic aggregates of the past period, plus some sparse and imperfect information of the current period. Non-extensive entropy technique has proved to remain a good modelling device not only in the case of high frequency series, but also in the case of aggregated series. To predict the missing GDP components, we set up a q-generalized Kullback–Leibler information divergence criterion function with a priori consistency, GDP related macroeconomic constraints and regular conditions. The model forecasts are compared to the official Polish GDP components of the corresponding period. The proposed Tsallis entropy approach leads to high predictive performance and shows a stronger estimation stability through different model simulations than the traditional Shannon model. Furthermore, as expected this Tsallis related approach seems to reflect a higher stability through parameter computation and simulation in comparison with the traditional Shannon–Gibbs entropy technique.

1. Introduction

This paper aims at extending applications of the non-extensive entropy approach through solving a generalized inverse problems of forecasting the components of gross domestic product(GDP) as stochastic macroeconomic systems. Highly competitive and technological age requires fast access to information. Trying to meet these needs, official statistics in particular prepares flash estimates which are published as soon as possible ($t + 45$, $t+30$, $t+15$). Next, immediately after each year/quarter period, it is a worldwide generalized practice to forecast the components of GDP for the current or/and future periods on a basis of existing series of the past periods, plus some sparse and incomplete information of the period under forecast. Since the beginning of the Cowles commission at the early of 1930, forecasting socio-economical events trough econometric techniques has remained a challenge. One of the existing methodology for econometric modelling and forecasting were developed, among others, by Tinbergen* [1] and Klein† [2] in applied dynamic models for economic processes and forecast. This approach, earlier developed by Marshal relies upon the analysis principle of "ceteris paribus" and suggests that economic phenomena are only partially connected. In reverse, the second school of modelling methodology, initially proposed by Warlas [3], relies on the general equilibrium principle, suggesting that socio-economic phenomena are fully interconnected and should not be analyzed separately. Both econometric methodologies rarely perform well once implemented for forecasting purposes. One of the reason of such a poor outcome is a bad knowledge of the data generating system. To simplify the problem, empirical econometric modelers apply, thanks to the central limit theorem, the Gaussian hypothesis with the risk, in many cases, of getting wrong or less stable solutions. A set of recent studies (e.g. [4–10]) seem to underscore that a large array of socio-economic phenomena take the form of a power-law (PL), in particular distribution of income, wealth, size of firms, macroeconomic scaling laws and distribution of financial variables.

Authors [11] have studied the dynamics of a general system of interacting units, each with a complex internal structure with many subunits which grow in a multiplicative way over a period of twenty years. They found that such systems display a PL. It is worthwhile to underscore the similarity of such a system with the internal mechanics of macroeconomic system, also composed of interacting economic sectors, each with a complex internal structure defined by firms exercising similar business. This list of power-law evidence in social science is not exhaustive. Following the above and taking into account that Gaussian model may be a limiting case of PL related Tsallis entropy model, we propose a q-generalized Kullback–Leibler information divergence (q-GID)‡ approach. There exists a few approaches for estimating GDP components. We apply an expenditure approach which is private, public and foreign sector oriented. According to that approach, the balanced equation, defining

---

GDP with respect to its components, is as follows:

$$\text{GDP} = \text{C} + \text{NPISH} + \text{G} + \text{GFCF} + \text{ALDV} + \text{INV}$$

$$+ X - M. \qquad (1)$$

where C is final consumption expenditures of households, NPISH is final consumption expenditures of non-profit institutions serving households, Gis final consumption expenditures of general government sector, GFCF is gross fixed capital formation, ALDV is acquisitions less disposals of valuables, INV is changes in inventories, X is exports and M is imports.

As shown in Table I, the expression GDP at the last row represents the quarterly GDP as a sum of its quartile components. The above expressions in parenthesis are later used in tables to explain the GDP components. Finally, as presented in the outputs paragraph, we shall use the official GDP component values from the Central Statistical Office of Poland (GUS) to gauge the performance of our proposed modelling technique.

2. Modelling generalized Ill-behaved inverse problems

2.1. Statistical problem setting

We forecast the national GDP quarterly components of the coming year on a basis of the GDP time series prior to this forecasted period to be adjusted by some incomplete or sparse data from the current forecast period. This information generally comes from the sampling surveys prior or related to the forecast period and is used to adjust the prior GDP components of the previous periods. Based on the Polish case (Table I), the most recent prior period is that of 2013, while adjusting information comes from the forecast year of 2014. The yearly periods from 2007 to 2012 have respectively served as GDP component priors for simulation purposes. The incomplete information of forecasted period was provided by the exports and imports of goods, the average paid employment in total enterprise sector and the average monthly nominal gross wage and salary in enterprise sector. These have played the role of linear approximates to adjust the information content of the 2013 prior. We have considered, as a starting hypotheses, that the proportion of exports (imports) of goods in exports (imports) structure, which includes exports (imports) services in 2014, will be the same as in the 2013 year of which all data is available. The same linear relationship was applied to adjust the household expenditure by the average monthly nominal gross wage and by the average paid employment in enterprise sector for assessing the gross fixed capital formation. The adjusted government expenditure prior has been computed through a linear trend. For the remaining accounts, the values of the prior come from the 2013 GDP components account, with no change. The same principle has been applied for the computation of the other lagged periods for simulation purposes. All the 32 values contained in the table are more or less approximate and do not statistically reflect the unknown, true values to be estimated.

Furthermore, there is no way to infer about the parameters of the true GDP component structure of the forecast period since, in this experiment, we are supposed not to process related official data. Thus, we deal with a stochastic inverse problem and no econometric method exists to cope with this kind of problem. In essence, the canonical ill-posed inverse problem as the one we deal with in this paper can be formally presented as it follows:

$$X(\zeta) = \int_D g(Y)h(Y,\zeta)\,\mathrm{d}Y + b(\zeta), \qquad (2)$$

where X means the observed matrix of updated prior, e.g. the 2013 GDP components data, Y designates the unknown matrix of the true 2014 GDP components values to be forecasted, D defines the Hilbert support space of the model, g is the transformation kernel linking measures X and Y, b explains random errors.

This is a basic model which consists of solving an integral equation of the first kind. Following [12], inverse problem recovery finds application in various fields of science, particularly in context of optimal control theory. Among different techniques proposed for solving this type of problems, the Tikhonov related regularization theory [13] remains the most applied, besides the Gibbs–Shannon–Jaynes (e.g. [14]) maximum (cross) entropy principle and the ill-posed stationary first-order Markov process, in which the operator can be seen as a generalized transition matrix and the Markov states. The contribution of this paper consists in extending the application of the non-extensive cross entropy formalism to search for global regularity-consistent with the maximum (non-extensive) entropy principle, while yielding the smoothest reconstructions of the Polish forecast of the GDP component system in the spirit of the Jaynes model [15].

2.2. The q-generalized Kullback–Leibler information divergence model

We follow recent works applying the non-extensive entropy econometrics (e.g. [16]) and define a q-GID criterion function to forecast the GDP component distribution, as it follows:

$$M \in H_q(p\|p^0, r\|r^0, \mu\|\mu^0) \equiv$$

$$\alpha \sum p_{klm} \frac{(p_{klm}|p_{klm}^0)^{q-1} - 1}{q-1}$$

$$+\beta \sum r_{\bullet lj} \frac{(r_{\bullet lj}|r_{\bullet lj}^0)^{q-1} - 1}{q-1} + \dots$$

$$+\delta \sum \mu_{k\bullet s} \frac{(\mu_{k\bullet s}|\mu_{k\bullet s}^0)^{q-1} - 1}{q-1}. \qquad (3)$$

Subject to

$$C_{\bullet l} = \sum_k ([C_{\bullet l} P_{kl}] + e_{\bullet l}) =$$

$$\sum_k^K \left(\left[\sum_{m>2}^M C_{\bullet l} v_{klm}(p_{klm}^q) \right]' + \sum_{j=1\cdots J} r_{\bullet lj}^q z_{\bullet lj} \right), \qquad (4)$$

$$H_{k\bullet} = H_{k\bullet} + \omega_{k\bullet} = H_{k\bullet} + \sum_{s=1}^{S} \mu_{k\bullet s}^{q} v_{k\bullet s}, \qquad (5)$$

$$\sum_{k=1}^{K} \sum_{j>2...J} p_{klm} = 1, \quad \sum_{j>2...J}^{J} r_{\bullet lj} = 1,$$

$$\sum_{s>2...S}^{S} \mu_{k\bullet s} = 1 \qquad (6)$$

where $C_{\bullet l}$ means each total by column l (observed values by quarterly l, including unknown error), $H_{k\bullet}$ means each total by row (observed values by GDP component k), p_{kl} probabilistic structure of GDP components across quarterlies, "$\bullet$" means variable related on the total row or column, according to context.

There exist a few types of constraining forms defining expectations in the Tsallis statistics. In the above model we apply the Curado–Tsallis (C–T) constraints [17–19], the general form of which is as follows: $\langle y_p \rangle = \sum_i p_i^q y_i$. The real q stands for the Tsallis parameter whose value should be around unity, suggesting the PL limiting Gaussian case. Since each row and column total sums are known with errors, Eqs. (3) and (4) display random terms respectively equal to $e_{\bullet l}$ and $\omega_{k\bullet}$. These random terms, like the probabilities p_{kl}, are redefined by the last terms of the same equations in support space formulation. For more details on the support space definition, we send the reader to e.g. [20]. The above criterion function $H_q(p\|p^0, r\|r^0, \mu\|\mu^0)$ measures the relative non-extensive entropy generated by the divergence between the prior and the posterior hypotheses through the forecasting process. The weights α, β, δ, introduced in the above dual objective function, define the trade-off between two types of terms (Golan [20]) which may exercise a significant impact on the model optimal outputs through the Lagrange multipliers linking the constraining information with the objective function. Likewise, it is worthy to recall that adding new consistency macroeconomic relations to the constraining part of the above model (Eqs. (3)–(8)) should increase the quality of the forecast in conformity with the maximum entropy principle connected with the Bayesian optimal processing rule. Finally, as far as the above parameter confidence area is concerned, we send the readers to [16]. It has been shown in [21] that computed indexes fulfil the basic Fisher–Rao–Cramer information index properties, including symmetry, continuity, additivity and maximum.

3. Outputs and discussions

The model has been computed through the GAMS code and data which come from GUS database [22]. The presented below outputs have been derived from the non-seasonally adjusted time series (period 2001 q1 to 2013 q4). In spite of the fact outputs from the seasonally adjusted series will not be presented here, we found their forecast error coefficient very close to that from the original data. The seasonal adjustment was carried out through the TRAMO/SEATS technique [23]. Table I

displays the targeted Tsallis, post relative entropy outputs and Table II — the percentage deviation of the forecasted GDP component outputs from its official component values. These changes are synthesized through the q-GID variance equal to 4.71×10^{-5}. This stands for the deviation average between the GDP component distribution prior and posterior. In the all cases, we found, as expected, the Tsallis-q parameter was around unity. Likewise, the computed outputs from the Shannon–Gibbs model were similar enough to those from the Tsallis related technique. Next, aiming at gauging the stability of the non-extensive entropy technique, we have carried out further simulations in which the 2013 prior was replaced by the priors of six year period back. Each of these priors was updated by the same available information of the 2014 forecast period.

Figure 1 displays the q-GID variances explaining the deviation between the posteriors and the priors (red color line) or between the posteriors and the official data from GUS. For the computation of the posterior of each period, the corresponding prior has been introduced into the equation system (2)–(6) and the related q-GID variance was calculated. As expected, the closer the forecast period the posterior is, the lower is the information loss of the q-GID technique, with respect to the official data from GUS. This is not the case for the line linking the posteriors and the priors. It does not show a trend on the analyzed period 2007–2013. This suggests that the posteriors fit well with the corresponding priors. In reverse, these stand worse with time back in information recovering by the entropy model.

TABLE I

Post entropy GDP components for quartilles (posterior).

component	KW1	KW2	KW3	KW4	total
C	265175.7	257440.6	260017.2	251011.8	1033645.25
NPISH	3766.44	3775.118	4041.427	3585.088	15168.072
G	80039.17	80793.39	81919.94	82898.56	325651.062
GFCF	47588.84	68833.14	75089.4	124035.3	315546.684
INV	4092.967	234.212	1689.846	2320.367	8337.392
ALDV	370.855	197.416	389.027	356.755	1314.052
X	193624.4	199039.5	200581	207953.9	801198.767
M	186874.4	189971.8	197097.9	199705.3	773649.338
PKB kwart	407783.9	420341.5	426630	472456.5	

TABLE II

Percentage deviation of the forecasted GDP component outputs from its true official component values for quartilles.

component	KW1	KW2	KW3	KW4	total
C	−0.145	−0.691	−1.085	−2.012	−0.966
NPISH	−5.843	1.115	−8.381	−8.698	−5.31
G	−10.59	−4.865	−8.322	8.248	−3.251
GFCF	8.872	5.964	5.984	5.853	6.376
INV	−15.837	89.37	50.441	−15.775	25.23
ALDV	−804.523	−373.419	−806.823	−685.803	−668.451
X	1.201	0.722	1.325	1.616	1.222
M	1.01	1.528	−0.509	3.31	1.363
PKB kwart	0.6	2.234	0.249	4.459	

Fig. 1. *q*-generalised information divergence between the posterior and respectively the prior and the official data.

4. Concluding remarks

With all else being equal, the above carried out computations tend to indicate that the entropy technique is stable enough and could be reliable for solving this kind of inverse problems. In spite of relatively high forecast precision, we observe a bad performance in the case of the acquisitions less disposals of valuables (ALDV) and changes in inventories (INV) components. Adding in the model a new piece of a information, consistent with the properties of that component, it should significantly improve the quality of its prediction. This should be the main direction for this research issue in the coming future. Particular attention should be paid while predicting the GDP components reflecting smaller proportions in the GDP structure. In fact, the Tsallis entropy estimators also belong to the family of the Stein estimators [16], indicating that smaller probabilities are shrunk and higher probabilities dominate in the solution space. Adding more consistency related information to less influential components will enhance parameter precision and then allow for recovering their influence while dampening that of bigger events.

References

[1] J. Tinbergen, *Econometrica* **1**, 247 (1933).
[2] L.R. Klein, *An Essay on the Theory of Economic Prediction*, Markham, Chicago 1970.
[3] L. Walras, *Éléments d'économie politique pure, ou, Théorie de la richesse sociale, Elements of Pure Political Economy, or, Theory of Social Wealth*, L. Corbaz & Cie, Lausanne 1874 (in French).
[4] B.J. Mandelbrot, *Business* **36**, 394 (1963).
[5] J. Kwapień, S. Drożdż, *Phys. Rep.* **515**, 115 (2012).
[6] A. Dragulescu, V.M. Yakovenko, *Physica A* **299**, 213 (2001).
[7] V.M. Yakovenko, J.B. Rosser, *Rev. Mod. Phys.* **81**, 1703 (2009).
[8] M. Jagielski, R. Duczmal, R. Kutner, *Acta Phys. Pol. A* **127**, A-75 (2015).
[9] T. Lux, *Appl. Fin. Econ.* **6**, 463 (1996).
[10] R. Rak, S. Drożdż, J. Kwapień, *Physica A* **374**, 315 (2007).
[11] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, UK 2000.
[12] S. Bwanakare, *Acta Phys. Pol. A* **127**, A-13 (2015).
[13] A.N. Tikhonov, V.I. Arsenin, *Solutions of Ill-Conditioned Problems*, Wiley, New York 1977.
[14] A. Golan, J.M. Perloff, *J. Econom.* **107**, 195 (2002).
[15] E.T. Jaynes, *Probability Theory: The Logic of Science*, Washington University, Washington 1994.
[16] S. Bwanakare, *Entropy* **16**, 2713 (2014).
[17] R.C. Venkatesan, A. Plastino, *Phys. Lett. A* **376**, 3470 (2011).
[18] C. Tsallis, R.S. Mendes, A.R. Plastino, *Physica A* **261**, 534 (1998).
[19] S. Abe, G.B. Bagci, arXiv:cond-mat/0404253, April 2004.
[20] A. Golan, G. Judge, D. Miller, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Wiley, Chichester (England) 1996.
[21] C. Tsallis, *Introduction to Non-Extensive Statistical Mechanics: Approaching a Complex World*, Springer, Berlin 2009.
[22] GUS, Verified estimation of GNP for zears 2010–2014, (in Polish).
[23] A. Maravall, *Computat. Stat. Data Anal.* **50**, 2167 (2006).