Was the Financial Crisis of 2008 Forecastable?

Barnabás Ács
Senior Account Manager
Thomson Reuters, London
E-mail: acsbarnabas@gmail.com

The fall of Lehman Brothers in September 2008, surprised the financial markets around the globe. The financial collapse immediately afterwards hit everybody unexpectedly. This article discusses whether deeper analysis of the US macroeconomic data would have been able to give hints to traders about the approach of the crisis of 2008.

KEYWORDS:
Financial crisis.
Forecast.

The National Bureau of Economic Research (NBER) published on 1 December 2008 that the United States had been in the state of recession since December 2007. Hereby recession became official, 3 months after the collapse of Lehman Brothers and 5 months after the historic high of the West Texas Intermediate (WTI) benchmark. Then almost 8 months had elapsed since Bear Sterns was saved by JP Morgan. 2008 passed by with the S&P 500 Index falling almost by 41 percent.

Several authors in literature (*Faber* [2009], *Ritzholtz* [2009], *Morris* [2009]) state that the current financial crisis was built in the system, and several renowned economists have blamed the regulatory bodies, mainly the Federal Reserve System (FED), being far too inactive preventing it (*Fleckenstein–Sheenan* [2008]). Based on their opinion, the crisis was encrypted in the system not only since 2001, but since the 1970s, the start of the real estate market boom. However, it seems that its amplitude, depth and arrival have surprised the general public. The aim of this study is to examine whether the 2008 economic crisis was forecastable, by means of statistical tools and time series models, and if there were any visible signs in the economic databases noticeable to everyone, to clarify whether its suddenness was the result of global blindness or the "storm clouds" were only visible to a few "insiders" only.

The macroeconomic indicators published by the national statistical agencies and supranational organisations (IMF, OECD, World Bank) are available basically for free to all investors and decision makers. This is especially true for the macroeconomic indicators of the United States.

The analysis spectrum of this paper is between the first quarters of 1985 and 2010. In this given 25 years, the NBER, the ultimate reference for economic cycles, identified three macroeconomic downturns.

The Japanese Banking Crisis of 1990, followed by the Dotcom Crisis of 2000 that had started with the burst of the Internet equities bubble and bottomed after the terror attacks of September 2001, and finally the financial meltdown starting at the end of 2007, referred to as the Crisis of 2008.

The length of the data series and the number of historical events make the analysis of the typical economic characteristics possible. The question of this article is: Would analysing the time series and their interactions with naive data mining tools have made possible to sense the approach of the crisis, or do even the most accurate statistical tools fail to unveil the signs?

However, the non-profit character of the study must be emphasized. It does not aim to find the most secret, always profitable investment strategy and indicator constellation. Neither does it try to criticise the main theses of various macro economic theories.

The article, however, sets the objective of clarifying the role of the different macro indicators eagerly followed by investors that influence their trading decisions, the real interactions thereof, as well as their short and long term effects on one another. It aims to uncover non-realised anomalies. In a holistic view, the goal is to check whether the general attitude towards statistics and econometrics is true: only afterwards they are clever. Hence, I am going to test numerous hypotheses.

It is assumed that the output indicators, in consequence of self-fulfilling prophecies, are not behaving as results in statistical models but as root causes. Moreover, it is presumed that not the indicators followed most intensively by investors carry the maximum amount of information about the state of the economy. Nota bene, the role of these indicators can change during time.

It can be hypothesized that if not their own trend-change but the change of interactions between indicators, can indicate the start of a recession. It can be experienced in several cases that most of the regressing equations lose their stability in the times of crisis. Hence the paper strives to find indicator pairs that are either stable in crucial situations or behave the same way before each crisis.

It is probable that a recession is indicated later due to error correction mechanism among variables in dynamic relationships. Moreover, it is also supposable that the explosion or birth of the dynamic relationships among variables carries useful information for the trend changes of the variables indicating the recession.

The study is based on the discussion of the definitions, nature and characteristics of the crises (for example *Kindleberger* [1989], *Fisher* [1933], *Minsky* [1977], *Dewald* [1972]. The main reasons of the current crisis (for example *Faber* [2009], *Fleckenstein–Sheenan* [2008], *Harris* [2008], Morris [2008], *Ritzholtz* [2009], *Király–Nagy–Szabó* [2008]) are also incorporated in the hypotheses and the findings. It demonstrates the purposes of the applied econometric methodology of analysis, goes through the results (model estimations, hypothesis tests) and points to further examination and research possibilities. Based on my goals, the following hypotheses were verified.

– *First hypothesis*: The output indicators, in contrast with prior expectations, and probably due to the self-fulfilling prophecies, are not endogenous but exogenous variables.

– *Second hypothesis*: When examining particular variables in different variable categories, it is important to analyze others than the ones followed by market participants as it is not sure that the emphasized variable carries the most information. Besides, the highlighted role can change from one period to the other.

– *Third hypothesis*: It is not the trend of the macroeconomic indicators that might be able to indicate the start of a recession but the change in the interaction between them.

– *Fourth hypothesis*: It is presumable that due to the necessary error correction mechanism among variables in dynamic relationships, the start of a recession can be hidden or indicated later, as similar events tend to characterize upswing times.

– *Fifth hypothesis*: The explosion or birth of the dynamic relationships among variables carries useful information for the trend changes of the variables indicating the recession.

In order to prove my hypotheses, data mining and knowledge discovery were applied.

1. Variables included

The article is analysing 140 macroeconomic indicators of the US, issued on a quarterly basis covering a time span between 1985 and 2010. As the recession of 2008, in contrast to that of 1990, started undoubtedly in the USA, it wasn't necessary to consider incorporating the macroeconomic indicators of other countries.

These 140 pieces of indicators – 99 chosen and 41 derived – enable to cover the whole economic sphere of the US, starting from GDP through price and interest rate levels and production data to the number of filings for bankruptcy protection. Those variables were chosen that fulfilled at least one of the two following criteria:

– They are in spotlight, i.e. the financial markets pay attention to them. This principle is fulfilled if they are highlighted in the Thomson Reuters data bases.

– They help to cover all aspects of the US economy, in order to enable the identification of the possible latent indicators. This goal is based on the "Guide to Economic Indicators" edited by *The Economist* [2006].

The chosen indicators fall into nine categories (*The Economist* [2006]). These are: *1*. indicators of value added, *2*. employment indicators, *3*. fiscal indicators, *4*. consumption indicators, *5*. investment and savings indicators, *6*. indicators of industry and commerce, *7*. indicators of balance of payments, *8*. money and financial markets indicators, and *9*. indicators of prices and wages.

It was not intended to have the same number of elements in each category. To eliminate absolute values, units and the effect of inflation, 41 indicators are expressed as a ratio of either current or constant priced GDP (*Hajdu–Virág* [1993]).

2. The structure of the analysis

For the purpose of meeting my research goals, I analysed the formerly mentioned data series. Firstly, I examined whether the chosen variables are stationary, so I applied the Dickey-Fuller regression for each y_t variable:

$$\Delta y_t = \mu + \beta y_{t-1} + \alpha_1 \Delta y_{t-1} + \ldots + \alpha_r \Delta y_{t-r} + \varepsilon_t \ ,$$

where $\beta = \delta - 1$. The existence of unit root in the augmented Dickey-Fuller test is proven by accepting the null hypothesis of:

$$H_0 : \delta = 1 \qquad H_1 : \delta < 1 \ ,$$

that is

$$H_0 : \beta = 0 \qquad H_1 : \beta < 0 \, .$$

Hence, we are considering a time series to be stationary, if the null hypothesis is rejected. The τ-test was used as test statistic:

$$\tau_\beta = \hat{\beta} \Big/ \left(se\left(\hat{\beta}\right) \right) \ ,$$

where $\hat{\beta}$ is the estimation of β and $se\left(\hat{\beta}\right)$ is the standard error of the estimated coefficient. We are considering a time series being stationary if $\tau_\beta > \tau_{critical}$.

Differencing was applied for all the nonstationary variables until the given variable was deemed to be stationary. Hence the order of integration of each variable could be specified.

In the second step, Granger-causality tests were used to identify the relationships between each y_t time series.

Applied to stationary time series – $x_t \sim I(0)$ and $y_t \sim I(0)$ – according to the null hypothesis x does not Granger-cause y, if no better estimate can be given to y compared to the case when only the past values of y are analysed. Id est:

$$H_0 : MSE\left(\hat{y}_t \,|\, y_{t-1}, y_{t-2}, \ldots\right) = MSE\left(\hat{y}_t \,|\, y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\right)$$
$$H_1 : MSE\left(\hat{y}_t \,|\, y_{t-1}, y_{t-2}, \ldots\right) > MSE\left(\hat{y}_t \,|\, y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\right)$$

where MSE stands for the mean squared error and $\hat{y}_t$ marks the estimated values of y. Based on this assumption, the following regression equation was applied on all the possible combination of the same order integrated variable pairs (x_t, y_t):

$$\hat{y}_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_l y_{t-l} + \beta_1 x_{t-1} + \ldots + \beta_l x_{-l} + \varepsilon_t .$$

The hypothesis system can be altered in the following:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1 : \exists j, \beta_j \neq 0 .$$

The test of this hypothesis system is quite straightforward with the Wald-test described also by *Jones* [1986]. So:

$$F_{emp} = \frac{MSE\left(\hat{y}_t \left| y_{t-1}, y_{t-2}, \ldots\right.\right)}{MSE\left(\hat{y}_t \left| y_{t-1}, y_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots\right.\right)}$$

with $(2l, T-2l-1)$ degree of freedom.

The null hypothesis is rejected, that is, it is presumed that x does Granger-cause y, if $F_{emp} > F_{(2l,T-2l-1)}$.

In the case of first and second order integrated time series, the analyses follow the same idea. Only the hypothesis system, the regression equations applied for the variable pairs and the Wald-test are modified with the first or second order differentiation.

To be able to measure the interdependency of the chosen variables, a so-called *cause-reason matrix* is compiled. This is a table with rows containing exogenous variables (being 'cause' in the Granger-causality tests) and columns including endogenous variables (reasons).

In the *cause-reason matrix* the exogenous and endogenous codes refer to the identifier of the given variable in the database (1 to 140). The cells of the table marked with "X" indicate the relationships where the *F*-values of the Granger-causality Wald-tests were significant. The matrix reveals that personal consumption (variable 1, referred to as: 1) does Granger-cause private investment (2), however private investment (2) does not Granger-cause personal consumption (1).

Using this matrix, a so-called "causality carpet" can be made, hence the relationships between variables can be easily qualified and quantified. It reveals, for example, that the free cash flow of institutions (7) – in column – is Granger-caused by the added value of households and institutions (5) and new orders of industries (8). At

the same time, however, the free cash flow of institutions (7) – in row – does Granger-cause amongst both private investments (2) and new industrial orders (8). Hence, it can be stated that a feedback relationship exists between the free cash flow of institutions (7) and new industrial orders (8).

Cause-reason matrix of the variables

Cause of #?	Exogenous Code	Reason of #?										
		10	68	16	43	32	0	32	40	33	4	...
		Endogenous Code										
		1	2	3	4	5	6	7	8	9	10	...
64	*1*	–	X		X	X			X	X		
29	*2*		–			X						
7	*3*			–								
48	*4*		X		–	X						
15	*5*					–		X				
0	*6*						–					
38	*7*		X					–	X			
58	*8*		X		X	X		X	–	X		
41	*9*		X			X			X	–		
5	*10*										–	
...	...											–

Note. The "Reason of #?" row states how many variables do Granger-cause the given variable. The "Cause of #?" column identifies the number of variables the given variable is a Granger-cause of.

The cause-reason matrix shows that personal consumption (1) is the Granger-cause of 64 variables, and there are only 10 variables that do Granger-cause personal consumption (1).

The main goal is to be able to classify whether a particular variable behaves as an exogenous or endogenous variable. To this end, normally the Hausman-test is used (*Hausman* [1978]); however, its programming in EViews is not possible. That is why a simple rule of thumb is applied. Based on the numerosity of the Granger-causalities, the variables are classified into four categories.

Classifying variables based on Granger-causalities:[1]

– Exogenous (ex): The given variable does Granger-cause twice as many variables than it is the cause of, i.e. *"Cause of #"* $\geq 2^*$ *"Reason of #?"*

[1] In the list, # stands for "how many".

– Rather exogenous (r-ex): The former criterion is not met, however, the given variable does Granger-cause more variables than it is the cause of, for example: *"Reason of #?" < "Cause of #?"< 2* "Reason of #?"*

– Not applicable (n.a.) "Reason of #?" = "Cause of #?"

– Rather endogenous (*r*-en): The given variable is Granger-caused by more variables than it does Granger-cause, for example: *"Cause of #?" < "Reason of #?"< 2* "Cause of #?"*

– Endogenous (en): The given variable is Granger-caused by twice as many variables than it does Granger-cause, for example: *"Reason of #?" ≥ 2* "Cause of #?"*

Regarding the former classification, each variable can be identified as being exogenous or endogenous.

In the third step, the variable pairs showing significant Granger-causality were described by the equation:

$$y_t = \alpha + \beta x_t + u \ .$$

It is examined whether the different recession times cause structural breaks in the particular regression relationships. That is why each regression equation was tested through the Chow breakpoint test (*Chow* [1960]). Its main point is to examine whether the parameters of the regression equations of the sub-periods of a given period, identified by one or more break-points, differ from each other, that is:

$$\begin{aligned} y_t &= \alpha_1 + \beta_1 x_t + \varepsilon_1 \\ y_t &= \alpha_2 + \beta_2 x_t + \varepsilon_2 \end{aligned} .$$

The hypothesis system according to this:

$$\begin{aligned} H_0 &: \alpha_1 = \alpha_2, \beta_1 = \beta_2 \\ H_1 &: \exists\left(\alpha_1 \neq \alpha_2, \beta_1 \neq \beta_2\right) \end{aligned} .$$

A structural break is identified, if equality of either regression parameters is rejected. The following Wald-test is applied for the test. The *F*-statistic is based on the comparison of the restricted and unrestricted sums of squared residuals and in the simplest case involving a single breakpoint, is computed as:

$$F_{emp} = \frac{\left(\tilde{e}'\tilde{e} - \left(e_1'e_1 + e_2'e_2\right)\right)/k}{\left(e_1'e_1 + e_2'e_2\right)/\left(T - 2k\right)} \ ,$$

where $\tilde{\varepsilon}'\tilde{\varepsilon}$ is the restricted sum of squared residuals, $e_2'e_1$ is the sum of squared residuals from subsample before the break, $e_2'e_2$ is the sum of squared residuals from subsample after the break. T is the total number of observations and k is the number of parameters in the equation, in this case $k=2$. This formula can be generalized naturally to more than one breakpoint.

We consider the breakpoint to be significant, if $F_{emp} > F_{(k,T-2k)}$. There is a constraint regarding placing the breakpoint, the Chow-test can only be carried out, if the size of the subsamples generated by the breakpoint is greater than the number of the estimated parameters, that is $T_i > k$.

In order to be able to examine the effect of recessions on a continuous basis, by indexing the breakpoints, the break-point test is conducted for each of the quarters between Q1 1988 and Q4 2008 (conforming the $T_i > k$ constraint). This way 84 pieces of Chow-tests are conducted and 84 pieces of *F*-values are calculated, that are used to build a time series of *F*-s for all the regressions:

$$z_t = \left(F_{1988q1}, F_{1988q2}, \ldots, F_{2008q3}, F_{2008q4}\right).$$

The creation of *F* time series enables the followings:

– As the critical F-value is the same in all of the tests (in this case $F_{(k,T-2k)} = 3.09$), if all the breakpoints F-values stay under this critical level, all the breakpoints should be considered not significant, which means that the regression is considered to be stable for the whole period, free from any structural breaks.

– By charting the empirical *F*-values, the quarters, where structural breaks occur, can be identified easily. This happens when either of the *F*-values exceeds the critical *F*-value. This way the breaks in the regression relationships preceding a recession can be identified. As previously mentioned, the time series analysed contain three recession periods identified by NBER: Q3 1990 to Q1 1991, Q1 2001 to Q4 2001, and Q4 2007 onwards.

Variable pairs with a structural break before the recessions at least two times out of the three can be identified easily.

In the fourth step the former regression pairs were put under scrutiny with the help of the Johansen-test checking for cointegration. Its purpose is to determine whether a group of non-stationary series is cointegrated or not.

EViews uses an identification method so that the error correction term has a sample mean of zero. We identify the part inside the error correction term by regressing the cointegrating relations $\beta' y_t$ on a constant (and linear trend).

When testing cointegration, the analyzed periods were:

– Q1 1986 to Q1 2010 to see which variable pairs were cointegrated in the whole period.

– Q1 1986 to Q4 2007 to make the comparison with Q1 1986 to Q1 2010, and identify those variable pairs that are not affected by the recession at all.

– The so-called inter-recession bands to identify the variable pairs with similar cointegration parameters in all three periods (Q1 1986 to Q3 1990, Q2 1991 to Q1 2001, Q1 2002 to Q4 2007).

– Q1 2002 to Q1 2010 to identify the variable pairs with changed parameters after the burst of the dotcom bubble.

The aim is to identify those variable pairs, that were cointegrated with the same β' parameters for more periods.

3. Results

The analysis works with a data set of 140 macroeconomic indicators looking back 25 years, published quarterly. When analysing Granger-causality, the significance of 11 084 variable pairs was checked and 843 Granger pairs were categorized. When running the Chow-test, 303 576 *F*-values were calculated and charted in 3 614 graphs. Doing the individual visual analysis of these, 828 characteristic regressions were chosen.

When testing cointegration, 4 968 Johansen-tests were run and analysed in six different time periods.

As the analysis was based on a wide literature and conducted with the utmost care, it can be stated that the main tendencies could be identified, and all material factors were discovered.

In congruence with the presumptions, it was proved that the majority of the economic time series are non-stationary. Out of 140 variables, 102 turned out to be first order integrated, meaning that the result of the regressions run on the non-transformed form of these variables will be biased. That is, the traders running the classic regression and correlation tests on these variables without differencing them, will base their decisions on wrong results.

Among others, the Granger-causality analysis revealed that in contrast with expectations, economic indicators measuring added value are primarily exogenous variables. The GDP is in feedback relationship with both the financial sector and the monetary policy.

The employment indicators can be solidly considered endogenous variables, however, no common factor can be identified that does Granger-cause the employment variables.

In contrast to expectations, the fiscal indicators are strongly endogenous and their primary Granger-causes are the profit and output indicators of the non-financial sectors.

In the case of consumption variables, it turned out that energy consumption and crude oil consumption handled similarly previously have completely different characters, the former shows endogenous, while the latter exogenous characteristics.

As for the investment indicators, it was proven that the expectations are manifested in profit indicators since the profit indicators are exogenous and in contrast, the stock indicators are endogenous.

Among indicators of commerce, the housing ones are rather endogenous, whereas those of car sales have no clear characteristics. Common factors can be identified, as short term consumption variables Granger-cause housing indicators, while it is the long term consumption indicators that determine the car market.

Neither do the indicators of the current accounts carry a common character, nor do common factors play a part in determining the budget deficit and current account balance.

A surprising finding is that the FED fund rate and the monetary base are exogenous variables.

Price indicators behave, as expected, clearly as exogenous variables.

The Chow-tests discovered inter alia that the added value of households and institutions have the most stable regression relationships among the value added indicators. It was shown that only the aggregate GDP and the GDP created by the government are in stable relationship with inflation.

The regressions including the employment indicators are not stable, this is one of the most important findings of the Chow-tests. Out of the five employment indicators only the participation rate has classifiable regressional relationship with other variables. Another interesting fact to note is: the Granger-causality tests showed that employment indicators are clearly endogenous variables. However, solid relationships are only built when they bear exogenous roles.

Out of ten fiscal indicators, merely four show solid endogenous relationships, hence it is also proven that this indicator group is also lacking steady relationships. It has to be emphasized, however, that government consumption and investment form stable relationships with all of its exogenous variables. The solid relationship between the fiscal and monetary indicators stresses the presumption that economic policymakers of the US are relying on both tools.

It is shown that the disposable income to GDP is in steady relationships with all the indicator categories. The crude oil consumption, in contrast to the energy consumption, forms solid relationship only with commodities price index, a sign that oil consumption is price-flexible.

Based on the Chow-tests, it can be stated that the profit of the financial institutions builds up solid relationships with the participation rate and the factors affecting their economic environment. However the profit of non-financial institutions is influenced by the indicators of the broader economy, such as consumption expenditures, consumer price index, and industrial production.

In case of the housing indicators, the focus has to be put on the housing inventory. This variable is influenced by the profit of the financial institutions, the trade balance, and the EUR/USD exchange rate in a solid way. The EUR/USD exchange rate tends to strengthen if the US economy is slowing down, which is in congruence with increasing housing inventories, the housing price decrease, and the declining profits of the financial institutions. Based on this, it could be expected that trade balance and the EUR/USD exchange rate are also in steady relationships; however, these tests show the opposite in both relations.

The retail sales of new passenger cars proved to be a stable variable, as it has stable relationships with all of its influencing indicators.

It is stated that the relationship between the consumer price index and the consumption expenditures on durables always breaks at times of recessions.

The indicators of money supply (M0, M1, M2) are in lagging relationships with their exogenous variables, the relationships always become weak after the start of the recessions.

There is a stable feedback relationship between the S&P500 index and the profit of the financial institutions to the GDP. This relationship states that long term price appreciation is not possible without the health of the financial institutions.

Based on the cointegration tests, it can be stated that the added value of the non-financial institutions is cointegrated in a constant parameter manner with three other variables; the consumer expenditures on services, the consumer expenditures, and GDP.

The ceasing character of the cointegration is proven in the case of the following variable pairs: *1*. Manufacturing production and the gross value added by businesses; *2*. Industrial production and constant price GDP; *3*. Private business sector production and manufacturing production.

Manufacturers durables new orders to the GDP and the private business sector production demonstrate the strength of the current recession, hence the cointegration parameters of the cointegrations equations are the same between Q1 1986 and Q3 1990 as well as Q1 1999 and Q1 2001.

The cointegration throughout the whole Q1 1986 to Q1 2010 period (referred to as holistic cointegration) between the government investment to GDP and the personal consumption stresses the automatism of the fiscal policy.

In the followings, the results are compared with the hypotheses set up.

First hypothesis: *The output indicators, in contrast with prior expectations, probably due to the self fulfilling prophecies, are not endogenous but exogenous variables*.

The result of Granger-causality tests showed that surprisingly the output indicators are strongly exogenous and they have feedback relationships with both the financial sector and monetary policy. Based on these facts, the first hypothesis can be accepted.

The self-fulfilling prophecies seem to be true in the case of the monetary policy as well; hence both the monetary base and the FED fund rate show exogenous characteristics. Special attention should be devoted to the monetary base, as its relationships with other indicators are stable. It has to be stated, however, that fiscal policy is an endogenous variable, which could be worrisome for those who are expecting an immediate result from fiscal policy.

Second hypothesis: *When examining particular variables, not only the ones followed by market participants but also others shall be analyzed, as it is not sure that the emphasized variable is carrying the most information. Besides, the emphasized role can change from period to period*.

This hypothesis can be accepted as well, mainly based on the behaviour of the employment indicators. These variables carry endogenous characteristics; however, they do not have any common influencing variables as Granger-causes. Apart from that, their regression relationships are not stable. Out of five employment indicators, only the regression relationships of the participation rate can be characterized. Still, the market does not follow this indicator.

Similar revelations can be made regarding the consumption of crude oil and that of energy. Firstly, the Granger-character of the two variables is totally different – the former is rather exogenous, the latter is endogenous –, secondly, energy consumption is basically in solid relationship with most of its variable pairs, while crude oil consumption has only one stable relationship. Thirdly, energy consumption is cointegrated with numerous variables, whereas crude oil consumption shows no cointegration.

These findings are important, because it shows that the focus of the financial markets is on the wrong indicator – huge volatily is occurring after the weekly crude stocks report (proxy indicator for consumption), but attention is paid to the energy consumption reports. All the traders want to know the non-farm payroll figures as they are released, however, as it seems from the foregoing, it is not a correct indicator to follow.

Third hypothesis: *Not the trend of macroeconomic indicators but the change in their interaction is able to indicate the start of a recession.*

This hypothesis is partially confirmed. The characteristic breaks of the regression relationships are only confirmed in the case of a few indicators. Regarding the output indicators, merely the constant price GDP and the output of the non-financial institutions to the GDP regressions break in a characteristic way.

As far as the employment indicators are concerned, only the regression relationships of the participation rate break in several cases, nor the lagging style break is proven. Within the fiscal, consumption, investment and commerce variable categories, the regression relationships of only the following variables break in a characteristic way: governmental consumption and investment to GDP, consumption expenditures for transportation, private investment, and aggregated industrial production. Neither the current account balance, nor the money market indicator has characteristic regression breakpoints, and the same can be stated for the price indicators.

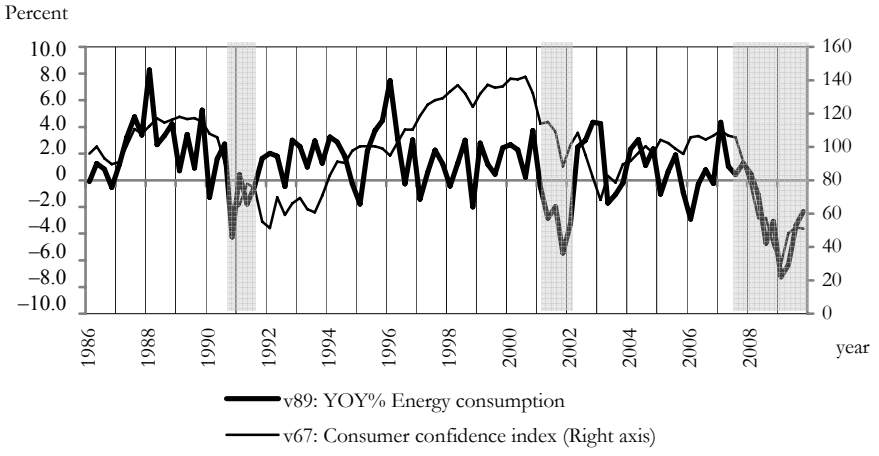
Fourth hypothesis: *It is presumable that, due to the necessary error correction mechanism among variables in dynamic relationships, the start of recession can be hidden or indicated later, and the same happens also in upswing times*.

This hypothesis is also partially proven, as only the holistic cointegration characteristics could cause these anomalies. However, holistic cointegration makes up only one third of the specific cointegration relationships, and it is mainly typical for consumption, investment, and commerce indicator groups.

Fifth hypothesis: *The explosion or birth of the dynamic relationships among variables carries useable information for the trend changes of the variables indicating recession*.

This hypothesis is clearly confirmed due to not only the numerosity of these cointegration relations but also the fact that starting and ceasing cointegrations are typical for specific indicators. Since the burst of the 'dotcom balloon', cointegrations ceased to exist in the case of the aggregated industrial production. However, the numerous cointegrations came to existence in the case of energy consumption and consumer satisfaction.

Energy Consumption and consumer confidence



Source: Thomson Reuters datastream.

In times of economic expansion, cointegrations that characterise the current price GDP, the output of non-financial institutes, energy consumption, the free cash flow of companies prognosticate new single family home sales and non-farm payroll costs.

Energy consumption proved to be a very important indicator of the state of the economy, as it is clearly endogenous in character, it has stable regression relationships, and carries numerous characteristic cointegrating relationships. Consumer satisfaction is to mention among exogenous variables. The importance of these two variables is enhanced by the fact that consumer satisfaction does Granger-cause energy consumption, the regression is stable, and since 2002 onwards these two variables has been cointegrated.

4. Conclusions

By using statistical methodology, the study identified variables and constellations that could have helped discovering the approach of an economic crisis. Based on the 'depth and width' of the database, it is presumable that only a fraction of the context was discovered.

The results support a further test on whether these variables can be 'wrapped into' latent variables (factors, main components) and the results can be generalised by means of these main components. However, it would need numerous changes in methodology, for example, in the dynamic factor models (*Tusnády–Zierman* [1987]).

Although my analysis was carried out at the macro level, it would be also worth analysing mezo- and micro-level data based on the usage of the same variables applicable to both industries and companies.

Since the credit crisis definitely started off from the US, this country was focused on. But it would be also important to check whether the same could be foreseen for other G7 economics.

References

CHOW, G. [1960]: Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*. Vol. 28. No. 3. pp. 591–609.

DEWALD, W. G. [1972]: The National Monetary Commission: A Look Back. *Journal of Money, Credit and Banking*. Vol. 4. Issue 4. pp. 930–956.

DICKEY, D. A. – FULLER, W. A. [1979]: Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*. Vol. 74. Issue 366. pp. 427–431.

FABER, D. [2009]: *And Then the Roof Caved in*: *How Wall Street's Greed and Stupidity Brought Capitalism to Its Knees*. John Wiley & Sons. Hoboken.

FISHER, I. [1933]: The Debt Deflation Theory of the Great Depression. *Econometrica*. Vol. 1. No. 4. pp. 537–557.

FLECKENSTEIN, W. A. – SHEEHAN, F. [2008]: *Greenspan's Bubbles*: *The Age of Ignorance at the Federal Reserve*. McGraw-Hill. New York.

HAJDU, O. – HERMAN, S. – PINTÉR, J. – RAPPAI, G. – RÉDEY, K. [1994]: *Statisztika I-II*. JPTE Kiadó. Pécs.

HAJDU, O. – VIRÁG, M. [1993]: Pénzügyi viszonyszámokon alapuló vállalatminősítés többváltozós statisztikai módszerek felhasználásával. *Ipargazdaság*. Vol. 44. No. 7. pp. 23–32.

HARRIS, E. S. [2008]: *Ben Bernanke's Fed: The Federal Reserve After Greenspan*. Harvard Business Press. Boston.

HAUSMAN, J. A. [1978]: Specification Tests in Econometrics. *Econometrica*. Vol. 46. No. 6. pp. 1251–1271.

HUNYADI, L. [2004]: Wald-próba a regresszióban. *Statisztikai Szemle*. No. 9. Vol. 82. pp. 866–869.

JONES, J. D. [1986]: Consumer Prices, Wholesale Prices, and Causality. *Empirical Economics*. Vol. 11. Issue 1. pp. 41–55.

KINDLEBERGER, C. P. [1989]: *Manias, Panics and Crashes: History of Financial Crises*. Basic Books. New York.

KIRÁLY, J. – NAGY, M. – SZABÓ, E. V. [2008]: Egy különleges eseménysorozat elemzése – a másodrendű jelzáloghitel-piaci válság és (hazai) következményei. *Közgazdasági Szemle*. Vol. LV. No. July–August. pp. 573–621.

MINSKY, H. [1977]: A Theory of Systematic Fragility. In: *Altman*, *E. I. – Sametz*, *A. W.* (eds.): *Financial Crises*: *Institutions and Markets in a Fragile Environment*. Wiley International. New York.

MORRIS, C. R. [2008]: *The Trillion Dollar Meltdown*: *Easy Money, High Rollers, and the Great Credit Crash*. Public Affairs. New York.

ODED, M. – ROKACH, L. [2005]: *Data Mining and Knowledge Discovery Handbook*. Springer. New York.

RAPPAI, G. [2010]: A statisztikai modellezés filozófiája. *Statisztikai Szemle*. Vol. 88. No. 2. pp. 121–141.

RITZHOLZ, B. [2009]: *Bailout Nation*: *How Greed and Easy Money Corrupted Wall Street and Shook the World Economy*. John Wiley & Sons. Hoboken.

SIPOS, B. [1986]: A Kondratyev-ciklus empirikus vizsgálata és prognosztizálása. *Statisztikai Szemle*. Vol. 64. No. 12. pp. 1209–1237.

SOREN, J. [1995]: *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press. Oxford.

THE ECONOMIST [2006]: *Guide to Economic Indicators*: *Making Sense of Economic*. Sixth Edition. Profile Books. London.

TUSNÁDY, G – ZIERMANN, M. [1987]: *Idősorok analízise*. Műszaki Kiadó. Budapest.