

Раздел 3.

Построение и анализ эконометрической модели (аналитический)

При построении эконометрической модели нами были использованы статистические данные файла, представленного в пакете Eviews, содержащего временные ряды реальных показателей за достаточно большой промежуток времени. Файл содержит квартальные данные для указанных показателей с 1947 года по 1994 год включительно:

Временной ряд	Обозначение
Personal consumption expenditures (bil. 1987\$)	CS
Use of gross savings: gross investment	INV
Gross domestic product (bil. 1987\$)	GDP

Так же в работе будем использовать лаги, т.е. значения временных рядов в предыдущие моменты времени, обозначая, например

Временной ряд	Обозначение
Gross domestic product (ВВП) в предыдущем квартале	GDP(-1)

Приведем основные определения и формулы, касающиеся непосредственно темы нашей работы и ее целей.

При анализе адекватности регрессионных моделей, наиболее важным является анализ случайных отклонений моделей на нарушение предпосылок метода наименьших квадратов, чтобы параметры являлись наиболее эффективными оценками в классе линейных несмещенных оценок или BLUE оценками. Для временных рядов используется так же понятие так называемого «белого шума».

Для проверки того, что остатки моделей являются «белым шумом» требуется, чтобы:

1. Случайные отклонения $\{e_i\}$ должны быть распределены согласно закону нормального распределения $\sim N(0, \sigma)$ (статистика JB).
2. Случайные отклонения e_i и e_j являются независимыми друг от друга для всех $i \neq j$.
3. Дисперсия случайных отклонений должна быть постоянной, т.е. случайные отклонения модели — гомоскедастичны.
4. Ряд случайных отклонений модели $\{e_i\}$, не только удовлетворяет сформулированным выше условиям (Гаусса-Маркова), но и является стационарным.

Целью нашей работы, очевидно, является проверка второго условия из перечисленных. В случае проверки случайных отклонений на автокорреляцию используются традиционные методы: графический анализ, метод рядов (Сведа-Эйзенхарта), статистика Дарбина-Уотсона, хотя для анализа временных рядов в числе наиболее эффективных и объективных можно назвать анализ автокорреляционных функций, Q-статистику и тест Бреуша-Годфрея.

Q-статистика в интерпретации Льюнга и Бокса (Ljung-Box) $LB = n(n+2) \sum_{i=1}^m \left(\frac{\rho_k^2}{n-k} \right)$

Выводы относительно присутствия корреляции будем делать по значениям соответствующих P-вероятностей, получаемых при сравнении значений Q-статистики с соответствующими критическими точками. В коррелограмме, которую будем выводить с помощью эконометрического пакета Eviews, представлены как значения самой статистики Льюнга-Бокса, так и P-вероятности, трактовать которые мы будем. В качестве нулевой гипотезы используется предположение об отсутствии корреляции на соответствующем лаге, для которого и смотрится, и трактуется значениями P-вероятности.

В тесте Бреуша-Годфрея строится регрессионная модель случайных отклонений построенной модели на константу, на объясняющие переменные построенной регрессионной модели и на распределенные лаги случайных отклонений. Количество вводимых в модель лагов отклонений определяется порядок корреляции. Например, если нам оценивалась модель: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, то для проверки серийной корреляции (автокорреляции второго порядка) случайных отклонений, строится модель $\varepsilon_t = \gamma_0 + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \delta_t$. В качестве нулевой гипотезы так же используется предположение об отсутствии корреляции случайных отклонений, а для принятия решения используется либо F-статистика для проверки совокупной статистической значимости введенных лагов отклонений, либо величина nR^2 , сравниваемая с точкой распределения $\chi^2(k)$, где k — количество лагов, выбранное для проверки теста Бреуша-Годфрея. Так же можно ориентироваться на значения P-вероятностей для коэффициентов лагов остатков во вспомогательной модели. Зависимости между указанными показателями существуют.

Прежде чем приступить непосредственно к построению моделей и их анализу, требуется проверить каждый из временных рядов наших переменных для определения их стационарности, и в случае нестационарности — для определения порядка интегрированности. От этого зависит, сможем ли мы построить по нашим временным рядам регрессионную модель традиционным методом наименьших квадратов без каких либо преобразований. Это возможно в том случае, если все показатели представлены стационарными временными рядами или если временные ряды являются коинтегрированными. Начнем проверку временных рядов на стационарность с визуализации. В нашем случае все временные ряды явно нестационарны, следовательно следующим встает вопрос об их коинтеграции. В основе понятия коинтеграции лежит идея о том, что в некоторых случаях отсутствие стационарности у многомерного процесса вызывается общим трендом, который может быть устранен, в результате чего эта линейная комбинация будет стационарной. При анализе графиков исследуемых временных рядов, следует обращать внимание на временные ряды, графики которых указывают на их нестационарность и к тому же обладают похожими трендами, что как раз наблюдается в нашем случае.

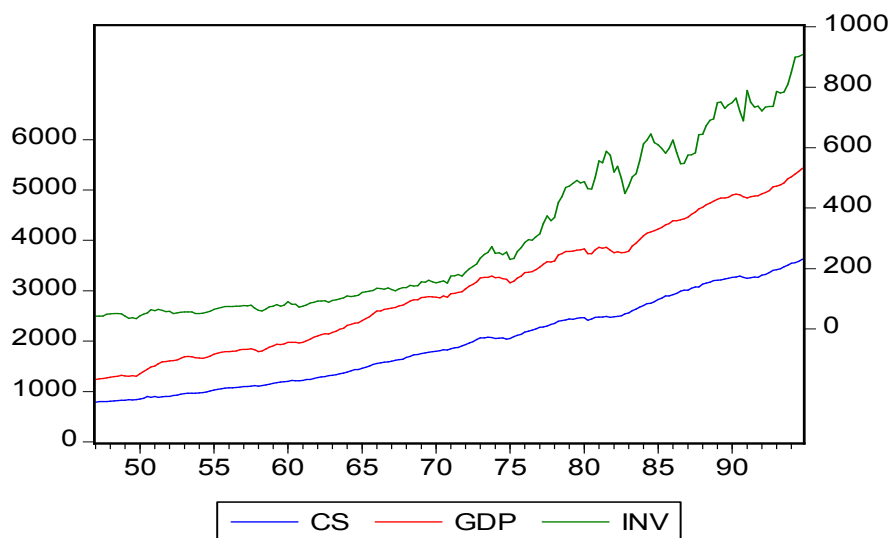


Рис. 1. Динамика рассмотренных временных рядов показателей CS, GDP, INV.

Correlogram of GDP

Date: 12/17/07 Time: 00:37						
Sample: 1947Q1 1994Q4						
Included observations: 192						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.983	0.983	188.57	0.000
		2	0.967	-0.005	371.78	0.000
		3	0.950	-0.007	549.73	0.000
		4	0.934	-0.003	722.53	0.000
		5	0.918	-0.005	890.27	0.000
		6	0.902	0.005	1053.2	0.000
		7	0.886	-0.007	1211.3	0.000
		8	0.871	-0.006	1364.7	0.000

Correlogram of D(GDP)

Date: 12/17/07 Time: 00:26						
Sample: 1947Q1 1994Q4						
Included observations: 191						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.359	0.359	25.032	0.000
		2	0.216	0.100	34.145	0.000
		3	0.076	-0.034	35.282	0.000
		4	0.054	0.021	35.867	0.000
		5	-0.011	-0.041	35.890	0.000
		6	0.018	0.030	35.954	0.000
		7	-0.057	-0.072	36.608	0.000
		8	-0.180	-0.172	43.147	0.000

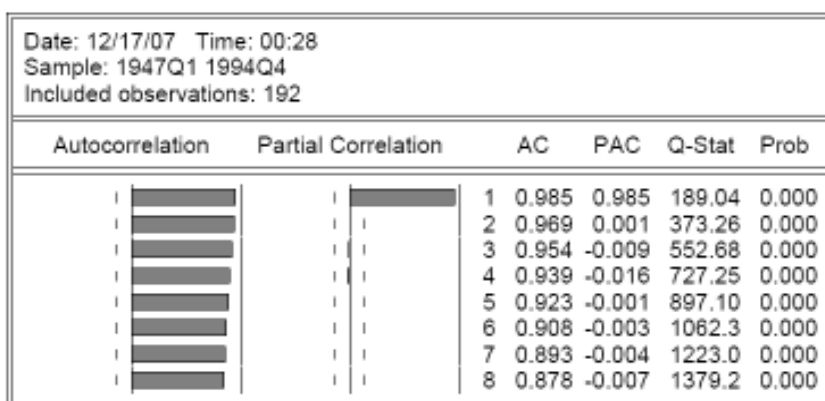
Correlogram of INV

Date: 12/17/07 Time: 00:27						
Sample: 1947Q1 1994Q4						
Included observations: 192						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.981	0.981	187.54	0.000
		2	0.961	-0.018	368.58	0.000
		3	0.941	-0.007	543.24	0.000
		4	0.923	0.018	712.01	0.000
		5	0.907	0.047	875.76	0.000
		6	0.892	0.021	1035.0	0.000
		7	0.877	0.008	1189.8	0.000
		8	0.863	0.003	1340.5	0.000

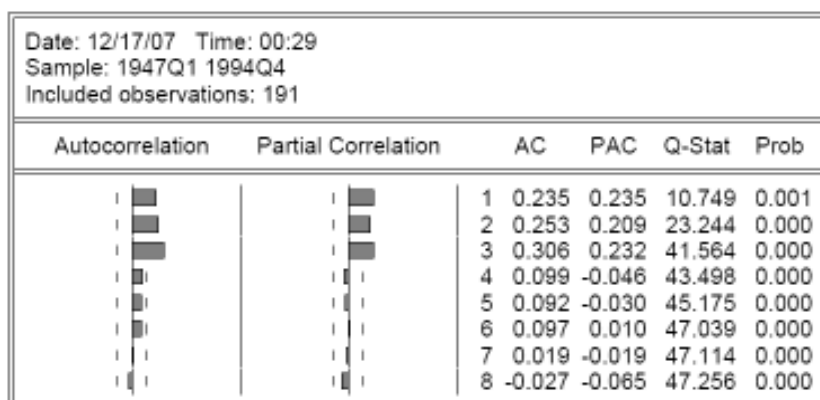
Correlogram of D(INV)

Date: 12/17/07 Time: 00:27						
Sample: 1947Q1 1994Q4						
Included observations: 192						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.060	0.060	0.6958	0.404
		2	-0.061	-0.065	1.4302	0.489
		3	0.199	0.208	9.1998	0.027
		4	-0.107	-0.147	11.450	0.022
		5	-0.036	0.017	11.705	0.039
		6	-0.141	-0.216	15.661	0.016
		7	-0.160	-0.080	20.797	0.004
		8	0.106	0.095	23.073	0.003

Correlogram of CS



Correlogram of D(CS)



Исходя из анализа коррелограмм наших временных рядов показателей, все ряды однозначно определяются как нестационарные временные ряды (присутствует угасающая экспонента на ACF и значимый первый лаг на PACF), стационарные в первых разностях, т.к. на соответствующих коррелограммах отсутствуют признаки, как нестационарности, так и принадлежности рядов первых разностей к классу TS (т.е. нет признаков содержания тренда). Это же подтверждается результатами тестов «единичного корня».

Временной ряд	Расширенный тест Дики-Фуллера			Тест Квятковского-Филлипса-Шмидта-Шина			Вывод
	Спецификация	ADF-стат	Крит. точки	Спецификация	LM-стат	Крит. точки	
CS	T,3	-1,611	-3,434	T	0,375	0,146	I(1)
Δ CS	T,2	-5,286	-3,434	T	0,036	0,146	I(0)
GDP	N,1	5,904	-1,943	T	0,329	0,146	I(1)
Δ GDP	C,0	-9,301	-2,877	T	0,029	0,146	I(0)
INV	T,0	-1,253	-3,433	T	0,442	0,146	I(1)
Δ INV	T,0	-13,305	-3,433	T	0,025	0,146	I(0)

Таким образом, на основе проведенного анализа принимаем решение о нестационарности рассматриваемых временных рядов. Для первых временных рядов в первых разностях, гипотеза о наличии единичного корня должна быть отвергнута, и, следовательно, трансформированные ряды являются стационарными, а исходные ряды – интегрированными первого порядка. Можно рассматривать вопрос о возможной коинтеграции временных рядов показателей CS, GDP, INV.

Базовая регрессионная модель, она же модель коинтеграции по Энглу-Грейнджеру, которую мы будем тестировать на наличие корреляции случайных отклонений, может быть записана следующим образом

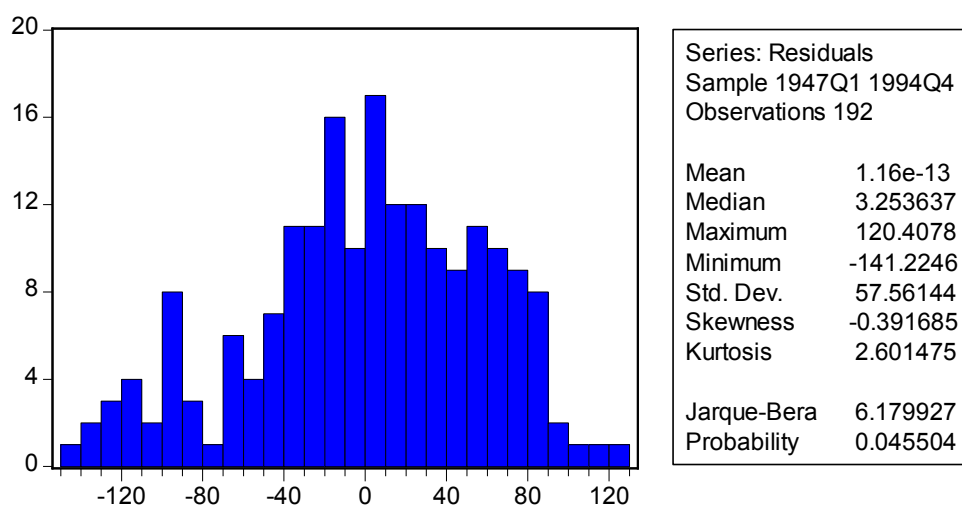
$$\text{GDP} = 166.7193428 + 1.535902443 \cdot \text{CS} - 0.4110576832 \cdot \text{INV}$$

В таблице приведены характеристики построенной модели, согласно которых все переменные, включая и константу, статистически значимы, но случайные отклонения при этом гетероскедастичны, коррелированы и не обладают нормальным распределением на уровне значимости $\alpha = 0,05$.

Dependent Variable: GDP
 Sample: 1947Q1 1994Q4
 Included observations: 192

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	166.7193	21.17921	7.871838	0.0000
CS	1.535902	0.020339	75.51512	0.0000
INV	-0.411058	0.066179	-6.211327	0.0000

R-squared	0.997713	Akaike info criterion	10.96961
Prob(F-statistic)	0.000000	Schwarz criterion	11.02051
Prob White (cross)	0.000000	Prob Breush-Godfrey (1)	0.000000
Prob White (no cross)	0.000003	Prob Breush-Godfrey (2)	0.000000
Prob Jarque-Bera	0.045504	Prob Breush-Godfrey (3)	0.000000



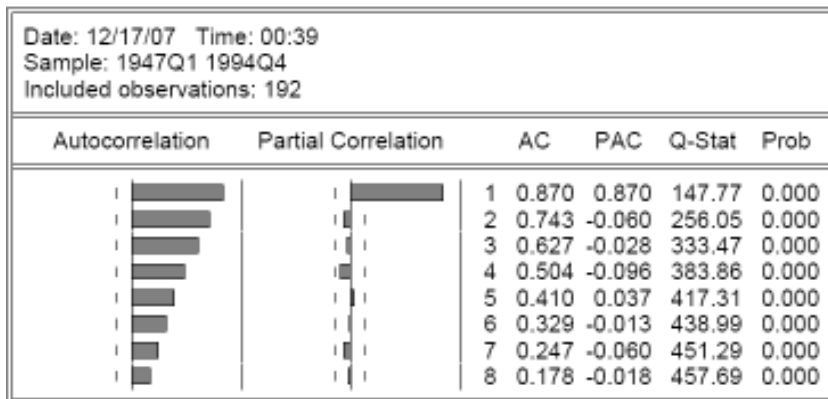
Значения асимметрии (Skewness), эксцесса (Kurtosis) достаточно близки к желаемым значениям, но статистика Жака-Бера и соответствующее ей P -значение 0,045504 говорят об отсутствии нормальности ошибок рассматриваемой модели при выбранном нами уровне значимости.

Для проверки наличия гетероскедастичности в остатках проведен тест Уайта, с регрессией квадратов отклонений модели на константу, регрессоры, их квадраты и попарные произведения. Результаты теста Уайта во всех случаях диагностируют гетероскедастичность случайных отклонений модели.

Для проверки наличия в модели автокорреляции случайных отклонений (автокорреляции первого порядка и серийной корреляции) используем LM -тест или тест Бреуша-Годфри для разного количества лагов, коррелограмму остатков, а так же делаем предварительный вывод как минимум об автокорреляции первого порядка для остатков – по значению статистики Дарбина-Уотсона 0,231837. Результаты LM теста Бреуша-Годфри, как и вид коррелограммы случайных отклонений, говорят о наличии автокорреляции остатков главным образом первого порядка (результаты в таблицах

ниже), не смотря на то, что в тесте Бреуша-Годфри гипотеза о коррелированности остатков принимается и для остальных значений лагов – сами лаги во вспомогательном уравнении теста незначимы. На коррелограмме присутствуют признаки коррелированности (для ACF характерна угасающая экспонента, но начинающаяся не с очень высоких значений, как у явно нестационарных временных рядов, P-вероятности для Q-статистики указывают на принятие гипотезы о коррелированности остатков).

Correlogram of RESID01



Временной ряд	Расширенный тест Дики-Фуллера			Тест Квятского-Филлипса-Шмидта-Шина			Вывод
	Спецификация	ADF-стат	Крит. точки	Спецификация	LM-стат	Крит. точки	
Resid	N,0	-3,647	-1,942	C	0,133	0,463	I(0)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	203.5928	Prob. F(3,186)	0.000000
Obs*R-squared	147.1795	Prob. Chi-Square(3)	0.000000

Test Equation:

Dependent Variable: RESID

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.66689	10.43188	1.118388	0.2648
CS	-0.013036	0.010061	-1.295741	0.1967
INV	0.046282	0.032843	1.409211	0.1604
RESID(-1)	0.916384	0.073173	12.52346	0.0000
RESID(-2)	-0.027407	0.099362	-0.275829	0.7830
RESID(-3)	-0.017162	0.073728	-0.232766	0.8162
R-squared	0.766560	Prob(F-statistic)	0.000000	
Adjusted R-squared	0.760285	Durbin-Watson stat	1.868347	

Случайные остатки регрессионной модели при этом являются стационарными, что подтверждает факт коинтегрированности временных рядов, не смотря на явные недостатки построенной модели.

Исходя из всего вышесказанного, попробуем улучшить нашу модель традиционными способами, а именно – введением лагов по соответствующим переменным, начиная с эндогенной (т.е. с использования авторегрессии). Введение лагов по отдельным переменным, включая переменную GDP в каждом случае порождало модель, случайные отклонения которой коррелированы, в итоге мы остановились на модели, в которой присутствовали лаги по всем объясняющим переменным.

$$\text{GDP} = 18.76482132 + 1.274191373 \cdot \text{CS} + 0.7707614432 \cdot \text{INV} + 0.9474946176 \cdot \text{GDP}(-1) - 1.201403194 \cdot \text{CS}(-1) - 0.7766854762 \cdot \text{INV}(-1)$$

Dependent Variable: GDP

Sample (adjusted): 1947Q2 1994Q4

Included observations: 191 after adjustments

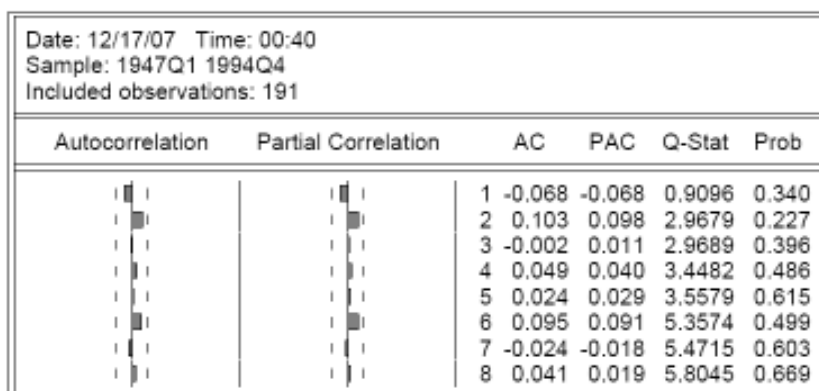
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	18.76482	6.691340	2.804345	0.0056
CS	1.274191	0.081907	15.55665	0.0000
INV	0.770761	0.064736	11.90622	0.0000
GDP(-1)	0.947495	0.020638	45.90938	0.0000
CS(-1)	-1.201403	0.093393	-12.86393	0.0000
INV(-1)	-0.776685	0.062536	-12.41984	0.0000
R-squared	0.999830	Akaike info criterion	8.392752	
Prob(F-statistic)	0.000000	Schwarz criterion	8.494917	
Prob White (cross)	0.000000	Durbin-Watson stat	2.126328	
Prob White (no cross)	0.000322	Prob Breush-Godfrey (1)	0.322609	
Prob Jarque-Bera	0.000000	Prob Breush-Godfrey (2)	0.248660	

Все переменные в модели, согласно своим значениям Р-вероятностей, статистически значимы. В построенной модели регрессии с лагами действительно решена поставленная перед нами задача – исправлена корреляция остатков предыдущей модели, так же в пользу этой модели свидетельствуют значения информационных критериев Акайке и Шварца, не смотря на то, что не решена проблема гетероскедастичности и возник вопрос с отсутствием у отклонений нормального распределения остатков. На самом деле, проблема кроется в известном уже факте – в длинном временном диапазоне, при сокращении которого, что достаточно легко проверяется с помощью пакета Eviews, по тем же временным рядам получают статистически значимые регрессионные модели, остатки которых гомоскедастичны и нормально распределены, следовательно, эта проблема может быть оставлена нами без внимания.

Таким образом, анализ коррелограммы остатков показывает, что остатки не просто не коррелированы, их коррелограмма в точности отвечает нашему представлению о коррелограмме «белого шума». Результаты теста Бреуша-Годфрея для любого количества лагов диагностируют отсутствие корреляции (как автокорреляции первого порядка, так и серийной корреляции для различного числа лагов в тесте). На коррелограмме Р-вероятности для Q-статистики свидетельствуют в пользу принятия нулевой гипотезы об отсутствии корреляции остатков. Автокорреляционные функции, в виду отсутствия на них значимых лагов, и уж тем более характерного рисунка – угасающей экспоненты или синусоиды, свидетельствуют о некоррелированности временного ряда случайных остатков и его стационарности. Случайные остатки регрессионной модели являются стационарными и согласно результатам тестов «единичного корня», что подтверждает факт коинтегрированности временных рядов, не смотря на явные недостатки построенной

модели. При определенных допущениях, сформулированных нами выше, модель может считаться статистически адекватной, раскрывающей цели нашей работы.

Correlogram of RESID02



Временной ряд	Расширенный тест Дики-Фуллера			Тест Квятского-Филлипса-Шмидта-Шина			Вывод
	Спецификация	ADF-стат	Крит. точки	Спецификация	LM-стат	Крит. точки	
Resid	N,0	-14,759	-1,943	C	0,073	0,463	I(0)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.908953	Prob. F(3,182)	0.437846
Obs*R-squared	2.819460	Prob. Chi-Square(3)	0.420306

Test Equation:

Dependent Variable: RESID

Variable	Coefficient	Std. Error	t-Statistic	Prob.
INV	-0.001881	0.065014	-0.028938	0.9769
CS	0.010541	0.082367	0.127978	0.8983
C	0.543947	6.896987	0.078867	0.9372
GDP(-1)	-0.002491	0.022930	-0.108637	0.9136
INV(-1)	0.001743	0.062632	0.027831	0.9778
CS(-1)	-0.006986	0.095679	-0.073013	0.9419
RESID(-1)	-0.060418	0.077713	-0.777451	0.4379
RESID(-2)	0.103747	0.077748	1.334401	0.1837
RESID(-3)	0.014446	0.077287	0.186912	0.8519
R-squared	0.014762	Mean dependent var	-1.51E-13	
Adjusted R-squared	-0.028546	S.D. dependent var	15.62150	
S.E. of regression	15.84290	Akaike info criterion	8.409294	
Sum squared resid	45681.52	Schwarz criterion	8.562542	
Log likelihood	-794.0875	F-statistic	0.340857	
Durbin-Watson stat	1.994025	Prob(F-statistic)	0.949000	

Учитывая то, что в модель входят лаги по всем переменным, а также с учетом знаков при этих лагах, соответствующая модель могла выглядеть и следующим образом:

$$D(\text{GDP}) = 0.06577726876 + 0.7698146378 * D(\text{INV}) + 1.235568839 * D(\text{CS})$$

Dependent Variable: D(GDP)
 Method: Least Squares
 Date: 12/16/07 Time: 23:03
 Sample (adjusted): 1947Q2 1994Q4
 Included observations: 191 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.065777	1.677712	0.039207	0.9688
D(INV)	0.769815	0.062593	12.29877	0.0000
D(CS)	1.235569	0.078497	15.74029	0.0000
R-squared	0.684287	Akaike info criterion		8.428443
Prob(F-statistic)	0.000000	Schwarz criterion		8.479525
Prob White (cross)	0.000000	Durbin-Watson stat		2.069782
Prob White (no cross)	0.000000	Prob Breush-Godfrey (1)		0.619607
Prob Jarque-Bera	0.000000	Prob Breush-Godfrey (2)		0.192782

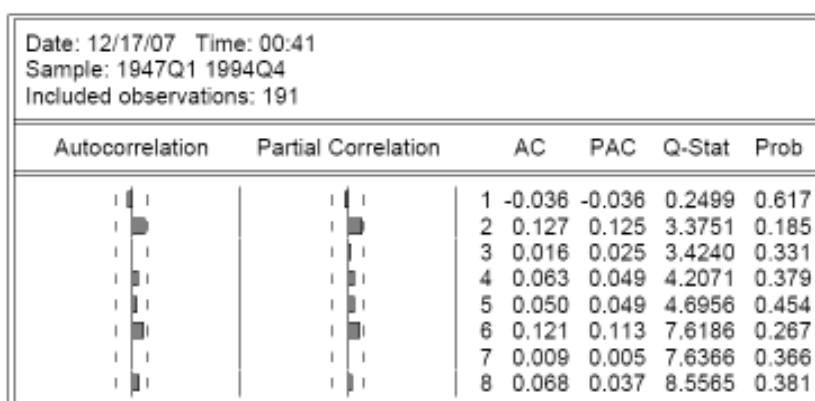
Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.122665	Prob. F(3,185)	0.341197
Obs*R-squared	3.415056	Prob. Chi-Square(3)	0.331947

Test Equation:
 Dependent Variable: RESID

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.176974	1.679171	-0.105394	0.9162
D(INV)	-0.000816	0.062584	-0.013031	0.9896
D(CS)	0.013372	0.078781	0.169732	0.8654
RESID(-1)	-0.034613	0.073562	-0.470529	0.6385
RESID(-2)	0.128231	0.073489	1.744910	0.0827
RESID(-3)	0.025699	0.073889	0.347809	0.7284
R-squared	0.017880	Mean dependent var	3.58E-15	
Adjusted R-squared	-0.008664	S.D. dependent var	16.15453	
S.E. of regression	16.22436	Akaike info criterion	8.441815	
Sum squared resid	48697.54	Schwarz criterion	8.543980	
Log likelihood	-800.1933	F-statistic	0.673599	
Durbin-Watson stat	2.004806	Prob(F-statistic)	0.643960	

Correlogram of RESID03



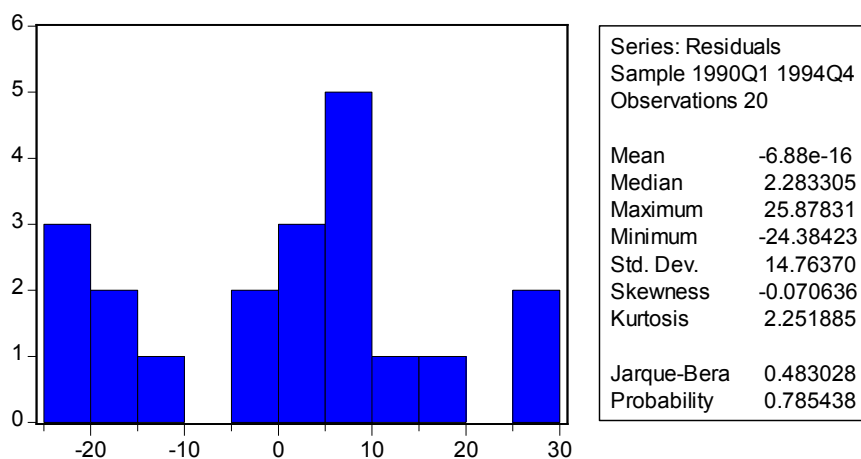
Статистические качества данной модели не отличаются от предыдущей и лишь подчеркивают наличие между рассмотренными временными рядами показателей не только долгосрочных, но и краткосрочных зависимостей, поскольку последняя модель – модель для приростов показателей. Случайные отклонения этой модели точно так же не коррелированы, стационарны (все результаты приведены в таблицах), но при этом гетероскедастичны и не имеют нормального распределения, для чего требуется формулировать допущения относительно временного промежутка, на примере последней модели покажем, что сокращение промежутка действительно корректирует гетероскедастичность и отсутствие нормального распределения остатков модели (в таблицах соответственно вывод характеристик модели, таблица статистики JB для определения нормальности распределения остатков и таблица теста Вайта).

Dependent Variable: D(GDP)

Sample: 1990Q1 1994Q4

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.460259	5.173958	0.088957	0.9302
D(INV)	0.222172	0.109427	2.030321	0.0583
D(CS)	1.369693	0.182383	7.509989	0.0000
R-squared	0.769399	Mean dependent var		28.85498
Adjusted R-squared	0.742269	S.D. dependent var		30.74430
S.E. of regression	15.60801	Akaike info criterion		8.470927
Sum squared resid	4141.372	Schwarz criterion		8.620287
Log likelihood	-81.70927	F-statistic		28.36018
Durbin-Watson stat	2.734409	Prob(F-statistic)		0.000004



White Heteroskedasticity Test:

F-statistic	1.526433	Prob. F(4,15)	0.244737
Obs*R-squared	5.785851	Prob. Chi-Square(4)	0.215722

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 12/16/07 Time: 23:30

Sample: 1990Q1 1994Q4

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27.86622	98.68861	0.282365	0.7815
D(INV)	-0.664102	2.193364	-0.302778	0.7662
(D(INV))^2	0.076933	0.040664	1.891899	0.0780
D(CS)	9.778103	5.089664	1.921169	0.0739
(D(CS))^2	-0.120957	0.119709	-1.010430	0.3283
R-squared	0.289293	Mean dependent var		207.0686
Adjusted R-squared	0.099771	S.D. dependent var		237.7030
S.E. of regression	225.5336	Akaike info criterion		13.88713
Sum squared resid	762981.1	Schwarz criterion		14.13607
Log likelihood	-133.8713	F-statistic		1.526433
Durbin-Watson stat	1.381474	Prob(F-statistic)		0.244737

Как и предполагалось, тест Вайта свидетельствует о гомоскедастичности случайных отклонений модели, а статистика Жака-Бера о их нормальном распределении, при этом статистическая значимость переменных модели присутствует (кроме константы).

На представленных данных продемонстрирована методика определения коинтегрированности временных рядов, включающая определение стационарности или нестационарности временных рядов, проверки принадлежности случайных отклонений моделей к «белому шуму». В работе удалось продемонстрировать не только, как отражается на результатах тестирования случайных остатков модели то, что исследователь может не учитывать некоторые свойства рассматриваемых временных рядов показателей, в данном случае свойство инерции, но и апробирован традиционный подход к исправлению в моделях положительной автокорреляции случайных отклонений. Показано, что для коинтегрированных временных рядов экономических показателей характерно наличие не только долгосрочных, но и краткосрочных взаимосвязей. Продемонстрировано использование коррелограммы и Q-статистики (трактовка ее P-вероятностей), а так же теста Бреуша-Годфрея при обосновании выводов о стационарности временных рядов показателей, а так же коррелированности случайных отклонений модели.