Gas pipeline infrastructure and economic growth of Russian regions: panel cointegration analysis

Alexander Belinsky[1,*]

[1]SC «Gazprom promgaz», 142702, Vidnoe, Moscow region, Russian Federation

Abstract. The Russian gas supply system is one of the largest infrastructures in the world. It is continuing to develop at a high rate. Modern methodological approaches and software allow to investigate the relationship between the development of energy infrastructure and economic dynamics of the regions. This paper proposes a methodological approach to the evaluation of the relationship "development of gas distribution systems – economic growth". The evaluation is based on the concept of cointegration and an error-correction models. This approach identifies both long-run and short-run components of this relationship. In this paper we aim to estimate the relationship using annual data of 17 Central Russia regions for the period 1998-2017. The results indicate that this relationship is statistically significant. It allows to apply proposed approach and the obtained estimates to the problems of analysis and forecasting of the domestic gas market.

1 Introduction

In recent years, there has been an intensive development of data processing techniques and software for empirical studies of the impact of infrastructure on the economy. However, gas supply infrastructure is rarely considered. According to conventional wisdom, the development of gasification leads to an increase in the investment attractiveness of the regions and has a positive impact on the economy. However, insufficient attention is paid to the relevant statistical studies.

In this paper we explore the statistical relationship "development of gas distribution systems – economic growth". Official statistics is presented in the form of time-series datasets, which are usually nonstationary and short, what requires the use of suitable specifications of mathematical models and methods for increasing the power of statistical tests. We investigate the hypothesis that the relationship is long-term, and the dynamics of the corresponding components has an equilibrium trajectory. The analysis is based on the concept of cointegration and the adjusting mechanism for panel data.

---

* Corresponding author: belinskyab@mail.ru

2 Overview

The development of the gas pipeline infrastructure of Russian regions is a large-scale socially significant project. The total investment since 2005 amounted to more than half a trillion rubles. Regulatory documents provide for the need to assess the impact of gas infrastructure development on economic growth, but in practice there are difficulties with carrying out these assessments. However, modern econometrics research offers many tools for time series studies and economic forecasting.

Among the first Aschauer empirically confirmed the existence of a close positive relationship between public investment and economic growth [6]. His works aroused great interest both in terms of opportunities to improve the validity of the state investment policy, and in terms of the development of the methodology of econometrics. A large number of follow-up works were devoted to specific issues – different regions of the world and different types of infrastructure were considered, new statistical methods were tested.

There are a lot of papers which are devoted to the impact of energy consumption on economic growth in various countries and regions of the world. The study of this relationship has become one of the classical applications of econometrics in energy researches. In [12] the authors studied 157 countries for the period 1960-2014 and confirmed the existence of long-term positive impact of electricity consumption on economic growth. Cointegration between Russia's GDP and electricity production is established in [2].

The results of the study [5] (67 countries for the period 1992-2005) showed that the increase in gas consumption by 1% in the long term could increase GDP by 0.652%. The estimation of the contribution of gas supply to the economic growth of 15 developing countries for the period 1990-2012, obtained in [4], is lower (0.069-0.102%). Researches of influence of development of gas distribution systems on economy of regions were conducted in the paper [3]. Analysis of data of 17 Russian regions for the period 1998-2015 demonstrated that the increase in gas consumption by 1% in the long term increases gross regional product (GRP) by 0.106-0.169%. But despite the differences in estimates, most of the papers confirm the significant impact of energy and gas supply on the economic growth of countries and regions of the world.

The methodology of modern research is based on the concept of cointegration, which is a statistical representation of the concept of economic equilibrium. The basic trend in research is to use longer time series panel and spatial data. Their availability, as well as the improvement of econometric tools, made it possible to improve the power of statistical tests and the statistical conclusion validity.

3 Methodology

Model. A basic approach to such studies provides for the use of an extended aggregate production function, where energy consumption is an independent production factor ([4, 5, 12, 20]). In addition, the choice of parameters should be consistent with the available information. We will consider "investment" form of the model, replacing capital on gross fixed capital formation.

Let $i = 1, ..., N$ be the sequence number of the region, $t = 1, ..., T$ be time period. The Cobb-Douglas formula will take the form:

$$Y_{it} = A_i e^{b_i t} D_{it}^\lambda I_{it}^\alpha L_{it}^\gamma Q_{it}^\beta V_{it}, \tag{1}$$

where Y_{it} – gross regional product (RUB million); L_{it} – average annual employment (thousand people); I_{it} – gross fixed capital formation (RUB million); Q_{it} – gas consumption (million m^3/year); D_{it} – length of gas distribution networks (km); A_i – technological parameter, which depends on the level of technology development (technical progress) and the development of

gasification; b_i – individual coefficients of productivity growth of the regional economy as a result of technical progress; α, γ, β – elasticity of output capital, labor and gas consumption, respectively; V_{it} – errors that are assumed to have zero mean and constant variance are not necessarily independent; λ determines a degree of influence of the gasification level on economic productivity.

Let's transform the expression given the scale (i.e. α+β+γ=1):

$$y_{it} = \mu_i + b_i t + \alpha i_{it} + \beta q_{it} + \lambda d_{it} + v_{it}, \qquad (2)$$

where $\mu_i = \ln(A_i)$ – fixed individual effects for regions; $y_{it} = \ln\left(Y_{it}/L_{it}\right)$; $i_{it} = \ln\left(I_{it}/L_{it}\right)$; $q_{it} = \ln\left(Q_{it}/L_{it}\right)$; $d_{it} = \ln(D_{it})$; $v_{it} = \ln(V_{it})$.

The model assumes that there is a long-term economic equilibrium between the considered parameters. The coefficients for variables are interpreted as elasticity. The choice of the main factors of economic dynamics is due, among other things, to an analysis of the initial publicly available information: statistical data from Rosstat and annually disclosed materials published by gas distribution organizations.

Cross-independence and unit root tests. The results of time series cross-independence testing provide the basis for choosing the reasonable type of unit root test. The choice of tests for cross-independence is due to the fact that the initial data is non-stationary, short and heterogeneous. So we apply tests [7,17].

The unit root tests are divided into the first generation tests (based on the assumption of independence between the rows in the panel) and the second generation tests (taking into account the existence of such a relationship). The tests of the first type include Maddala and Wu (1999), Tsoi (2001), Levin, Lin and Shu (2002), Im, Pesaran and Shin (2003), the second type – Demetrescu, Hassler and Tarcolea (2006), Hanck (2008), Pesaran (2007), [9,18], etc.

`**Cointegration Tests**. Pedroni [13, 14] proposed seven modifications of cointegration tests (based on the statistics υ, *rho, PP* and *ADF*): four of them are intragroup, three – intergroup. All tests might be applicable to tests of the null of no cointegration. The differences between panel and mid-panel panel tests consist in the specification of an alternative hypothesis. Also the *ADF*-statistics is a more powerful test in the case of small and medium sample size.

The second cointegration test, which is used in the present paper, is the test of Westerlund (2007) [19]. It is very general and allows to take into account the large heterogeneity in the regions included in the panel, both in terms of the parameters of the long-term cointegration equation and short-term effects of interaction between variables [15]. The test permits cross-sectional dependencies in the data and can use bootstrap techniques to calculate critical values of test statistics.

The null hypothesis of the Westerlund test [19] is the absence of cointegration. The alternative hypothesis depends on the specification of the test. The mean group tests (*Gt* and *Ga*) use the alternative hypothesis that there is cointegration for at least one region. Panel tests (*Pt* and *Pa*) use an alternative hypothesis that the panel is cointegrated as a whole.

Evaluation of the parameters of the long-run and short-run relationship. In the case when the considered variables have the same order of integration and the hypothesis of no cointegration was rejected, the further studies of the relationship "development of gas distribution systems – economic growth" is conducted using the model (2). To estimate the parameters of long-term relationship, we use: Fully Modified Ordinary Least Squares (*FMOLS*), Dynamic Ordinary Least Squares (*DOLS*).

To study the relationship of variables both in the long-term and in the short-term period, we apply Pooled Mean Group estimator (*PMG)* [16], using an Autoregressive Distributed Lag model *ARDL*(p, q_1, q_2, q_3). When using *PMG* estimates, we assume uniform equilibrium long-term dynamics of the considered variables in all regions.

An Error correction model (Equilibrium-correction model) is:

$$\Delta y_{it} = \mu_i + \phi^{ecm}\big(y_{i,t-1} - \alpha i_{it-1} - \beta q_{it-1} - \lambda d_{it-1}\big) + \sum_{j=1}^{p-1} \tilde{\varphi}_{ij}^{(y)}\Delta y_{i,t-j}$$
$$+ \sum_{j=0}^{q_1-1} \tilde{\varphi}_{ij}^{(i)}\Delta i_{i,t-j} + \sum_{j=0}^{q_2-1} \tilde{\varphi}_{ij}^{(q)}\Delta q_{i,t-j} + \sum_{j=0}^{q_3-1} \tilde{\varphi}_{ij}^{(d)}\Delta d_{i,t-j} + \epsilon_{it}, \tag{3}$$

where $i = 1, \dots, N$ – the sequence number of the region; $t = 1, \dots, T$ – time period; p, q_1, q_2, q_3 – selected model parameters $ARDL(p, q_1, q_2, q_3)$; α, β, λ – coefficient showing the long run relationship between variables; $\tilde{\varphi}_{ij}^{(\cdot)}$ – parameters of short-run effect of explanatory variables on the dependent variable; ϵ_{it} – errors with zero mean and finite variance; ϕ^{ecm} – coefficient characterizing the systems rate of return to the equilibrium state. In the presence of cointegration, the coefficient ϕ^{ecm} should be a statistically significant negative value. Modern software offers different criteria for selection of the specification of the model *ARDL*: Akaike information Criterion (*AIC*), Schwarz information criterion (*SIC*), Hannan-Quinn information criterion (*HQ*) [1]. We confirm the choice of *PMG* estimator using the Hausman test.

The Granger causality test. The concept of cointegration was proposed by Granger in 1981. We will also consider the Granger causality and will estimate the following models, including the error-correction mechanism.

$$\Delta y_{it} = \mu_{1i} + \sum_{k=1}^{p} \gamma_{11,k}\Delta y_{i,t-k} + \sum_{k=1}^{q} \gamma_{12,k}\Delta i_{i,t-k} + \sum_{k=1}^{q} \gamma_{13,k}\Delta q_{i,t-k} + \sum_{k=1}^{q} \gamma_{14,k}\Delta d_{i,t-k}$$
$$+ \phi_1 ECM_{i,t-1} + u_{1it} \tag{4a}$$

$$\Delta i_{it} = \mu_{2i} + \sum_{k=1}^{p} \gamma_{21,k}\Delta y_{i,t-k} + \sum_{k=1}^{q} \gamma_{22,k}\Delta i_{i,t-k} + \sum_{k=1}^{q} \gamma_{23,k}\Delta q_{i,t-k} + \sum_{k=1}^{q} \gamma_{24,k}\Delta d_{i,t-k}$$
$$+ \phi_2 ECM_{i,t-1} + u_{2it} \tag{4b}$$

$$\Delta q_{it} = \mu_{3i} + \sum_{k=1}^{p} \gamma_{31,k}\Delta y_{i,t-k} + \sum_{k=1}^{q} \gamma_{32,k}\Delta i_{i,t-k} + \sum_{k=1}^{q} \gamma_{33,k}\Delta q_{i,t-k} + \sum_{k=1}^{q} \gamma_{34,k}\Delta d_{i,t-k}$$
$$+ \phi_3 ECM_{i,t-1} + u_{3it} \tag{4c}$$

$$\Delta d_{it} = \mu_{4i} + \sum_{k=1}^{p} \gamma_{41,k}\Delta y_{i,t-k} + \sum_{k=1}^{q} \gamma_{42,k}\Delta i_{i,t-k} + \sum_{k=1}^{q} \gamma_{43,k}\Delta q_{i,t-k} + \sum_{k=1}^{q} \gamma_{44,k}\Delta d_{i,t-k}$$
$$+ \phi_4 ECM_{i,t-1} + u_{4it} \tag{4d}$$

where $\gamma_{\cdot\cdot,\cdot}$ and $\phi_{\cdot}$ – regression coefficients; $u_{\cdots}$ – errors with zero mean and finite variance; $ECM_{i,t-1}$ – the error-correction mechanism with using the two-step Engle and Granger procedure [10]. For the analysis of causality and exogeneity we estimate the coefficients $\gamma_{\cdot\cdot,\cdot}$ and $\phi_{\cdot}$. The corresponding variable is said to be weakly exogenous in this model, if the effect on it of the long-term dynamics of the variables is not statistically significant (i.e. the coefficient $\phi_{\cdot}$ is non-significant), but there is a short-term dependence on other variables (i.e. the coefficients $\gamma_{\cdot\cdot,\cdot}$ are significant). If in the long-term and in the short-term period the variable of interest does not depend on other variables, then it can be considered as a strictly exogenous variable in this model (i.e. the hypothesis indicating that the all coefficients are zero cannot be rejected).

To avoid the inconsistent estimators, we can use a system generalized method of moments (*GMM*) Blundell-Bond [8]. Also, Sargan statistics can be used to test for the validity of the instruments. In addition the identification of the parameters also requires no autocorrelation, which is tested using Arellano and Bond's Autoregressive (*AR*) test.

4 Empirical results

Our study uses annual time series for the period from 1998 to 2017 for 17 regions of Central Russia, published in open sources.

Tests Pesaran (*CD*-statistics) and Baltagi, Feng and Kao (*LM$_{BC}$*-statistics) reject the null of no cointegration for all variables at 0.01 significance level. Therefore, for the unit root tests we use the second generation tests, studying the model with a constant and a model with a constant and a trend. Note that the lag length for each regression is calculated automatically at the maximum length of 3. We concluded that all variables are assumed to be integrated of order one. This means that a long-term relationship between variables is possible, and it can be evaluated using the panel cointegration statistics.

Pedroni and Westerlund tests were performed to test for the null of no cointegration. Most of the tests rejected the respective null hypothesis. More powerful tests in the case of small and medium sample size rejected the hypothesis at 0.01 level.

The results allow to proceed to an assessment of the parameters of the long-term and short-term relationship (see table 1). The obtained estimates are statistically significant.

Table 1. Estimation of the long-run parameters

Variables	FMOLS	DOLS
i	0.421***	0.319***
q	0.115**	0.173***
d	0.131***	0.149***
R^2	0.81	0.90

Notes: Significance: *, **, *** – 0.1, 0.05, 0.01 level.

Thus, from table 1 it follows that in the long term the increase in the length of networks by 1% increases GRP by 0.131-0.149%, the growth of gas consumption by 1% increases GRP by 0.115-0.173%.

For the validity of using the *PMG* estimator to assess the parameters of short-term relationship, Hausman test was conducted, which did not reject the hypothesis of the validity of the method. The maximum lag length of the dependent and explanatory variables is set to 3.

Two specifications of the autoregressive model were studied in [3]. In this paper we selected *ARDL*(3,3,3,3) as the best performed *ARDL* specification. The obtained results are reported in table 2.

Table 2. Estimation of the parameters of *ARDL*(3,3,3,3) model

Variables	ARDL(3,3,3,3)
i	0.409***
q	0.086**
d	0.165***
Error-correction model	
ϕ^{ecm}	−0.299**

Notes: 1) Data is centered. 2) Significance: *, **, *** – 0.1, 0.05, 0.01 level.

The empirical results confirm the cointegrating relationship in the data, as well as a statistically significant positive impact of gasification on the economy in the long term. As for the investigation of short-term effect, in the obtained model the corresponding coefficients for all first differences are not significant. Therefore, it can be assumed that there is no short-term effect.

Exploration of endogeneity (exogeneity) was performed using *GMM*. Sargan test and Arellano and Bond's autoregressive test confirmed the validity of the instruments. In the long and the short term the rate of development of gas distribution systems doesn't depend on the GRP and the volume of gas consumption, i.e. is a strictly exogenous variable. But the volume

of gas consumption can be considered as a weakly exogenous variable due to the presence of short-term dependence on GRP and the rate of development of gas distribution systems.

5 Conclusions

The results of the study show that the application of the concept of cointegration and error correction mechanism allows to assess the relationship "development of gas distribution systems – economic growth". It is shown that the increase in the length of gas pipeline systems and the volume of gas consumption has a statistically significant positive impact on the economic growth of the regions. In the long term, the increase in the length of gas pipeline systems by 1% increases GRP by 0.131-0.149%, the growth of gas consumption by 1% increases GRP by 0.115-0.173%. The absence of a short-term effect can be assumed based on the analysis of the selected specification of the distributed lag autoregressive model. It is also argued that the length of gas distribution systems is a strictly exogenous variable and doesn't depend on the economic growth and the volume of gas consumption. But the volume of gas consumption can be considered as a weakly exogenous variable.

Such studies can be carried out for all regions of Russia. The results can be used to support decision-making in the planning of energy infrastructure development.

References

1. S.A. Ayvazyan, D. Fantazzini, *Econometrics-2: advanced course with applications in finance* (2017)
2. R.Y. Arkhipov, P.K. Katyshev, Applied econometrics, **44**, 38 (2016)
3. A.V. Belinsky, Gas industry, Special edition, **2** (2018)
4. S. Alam, S. Paramati, and other, Department of economics. Discussion paper, **21/15** (2015)
5. N. Apergis, J. Payne, Applied Energy, **87**, 2759 (2010)
6. D.A. Aschauer, Journal of Monetary Economics, **24**, 171 (1989)
7. B. Baltagi, Q. Feng, C. Kao, Journal of Econometrics, **170 (1)**, 164 (2012)
8. R. Blundell, S. Bond, Journal of Econometrics, **87(1)**, 115 (1998)
9. M. Costantini, C. Lupi, *A Simple Panel-CADF Test for Unit Roots, Economics Series*, **261** (Institute for Advanced Studies, Vienna, 2011)
10. R. Engle, C. Granger, Econometrica, **55**, 251 (1987)
11. C. Lee, C. *Chang,* Resource and Energy Economics, **30**, 50 (2008)
12. https://mpra.ub.uni-muenchen.de/79532
13. P. Pedroni, Panel Cointegration, Indiana University Working Paper in Economics (1997)
14. P. Pedroni, Oxford Bulletin of Economics and Statistics, **61,** 653 (1999)
15. D. Persyn, J. Westerlund, The Stata Journal, **8**, 232 (2008)
16. M. Pesaran, Y. Shin, R. Smith, J. of the American Statistical Association, **94**, 621 (1999)
17. M. Pesaran, CESifo Working, **1229** (2004)
18. M. Pesaran, Journal of Applied Econometrics, **22**, 265 (2007)
19. J. Westerlund, Oxford Bulletin of Economics and Statistics, **69**, 709 (2007)
20. P. Zweifel, P. Praktiknjo, G. Erdmann, *Energy Economics. Theory and Applications* (Springer International Publishing AG, 2017)