DOI: 10.5281/zenodo.22963172
Link: https://zenodo.org/records/22963172

EMPIRICAL ASSESSMENT OF THE INVERSE RELATIONSHIP BETWEEN ACTIVE BUSINESS DENSITY AND SHADOW ECONOMY SCALE ACROSS REGIONS OF UZBEKISTAN

Murodxujayeva Feruza Majidovna
Associate Professor, PhD, Namangan State University
Phone: +998999766675; E-mail: feruz_7881@mail.ru

Abstract – *Shadow economic activity in Uzbekistan differs sharply across territories, yet the association between formal business density and informality has received little regional econometric scrutiny. This study examines whether a larger stock of active enterprises accompanies a smaller shadow economy share in the 14 administrative territories of Uzbekistan. Official statistics on active enterprises, gross regional product (GRP) and the informal and hidden economy were merged into a balanced 2017–2025 panel of 126 observations, with 2024 serving as the reference cross-section. An ordinary least squares log-log model, ln(Shadow_Share) = β0 + β1 ln(Enterprises) + β2 ln(GRP) + ε, was estimated and tested for multicollinearity, heteroscedasticity, residual normality, functional form and influential observations, then re-estimated with year effects, two-way fixed effects and long differences. The 2024 model explained 68.8% of the variance (adjusted $R^2 = 0.632$; $F = 12.14$, $p = 0.002$). The GRP elasticity was −0.908 ($p < 0.001$): a 1% larger regional output corresponded to a 0.90% lower shadow share, negative in every specification and significant in all panel and long-difference estimates. The enterprise elasticity was positive (0.850, $p = 0.005$) once GRP was controlled, statistically equal in magnitude to the GRP elasticity, and vanished without Navoi and under fixed effects; the bivariate elasticity was −0.145 ($p = 0.484$). Formalization therefore tracks output per registered enterprise rather than enterprise numbers. Policy should therefore favor enterprise scaling, digital tax compliance and lower transaction costs, while recognizing that the very small sample, preliminary 2024 GRP data and partly accounting-based link limit causal interpretation.*

Keywords: *shadow economy, informal economy, business density, gross regional product, log-log regression, elasticity, formalization, regional heterogeneity, Uzbekistan, small and medium enterprises*

INTRODUCTION

The shadow economy, understood as market-based production of goods and services that are lawful in themselves but are concealed from public authorities in order to avoid taxes, social contributions or regulatory compliance, remains a persistent feature of economies at every level of income [1]. Cross-country estimates indicate that it is largest in low- and middle-income economies and recedes only gradually as development proceeds [2, 3]. Informality erodes the tax base, creates unequal competition between registered and unregistered producers, denies workers social protection and complicates the measurement of output [1, 4]. Its determinants, including fiscal burden, regulatory discretion, institutional quality and labor-market rigidity, have been studied extensively, although their relative weight differs across countries and estimation methods [5, 6, 7].

Uzbekistan offers an instructive setting for regional analysis. Since 2017 the country has

pursued a broad programme of economic transformation that included liberalization of the foreign-exchange regime, simplification of business registration and tax administration, and a larger role for the private sector in production. Official statistics reflect the scale of this change: the number of active enterprises and organizations rose from 268,428 in 2017 to 592,371 in 2023, nominal gross domestic product increased from 369.6 to 1,535.4 trillion soum between 2017 and 2024, and the official estimate of the informal and hidden economy fell from 45.6% to 34.3% of GDP over the same period. The National Statistics Committee defines the informal economy as production by households or individuals without established registration, and the hidden economy as activity that is not prohibited by law but is deliberately concealed from state bodies to avoid taxation and legal requirements. Following this official usage, the term shadow economy is used below for the combined informal and hidden economy.

National averages, however, conceal a sharp territorial divergence. In 2024 the shadow share of gross regional product ranged from 17.5% in Navoi region and 25.2% in Tashkent city to 53.0% in Namangan region, 54.9% in Khorezm region and 56.2% in Surkhandarya region, a spread of 38.7 percentage points. Tashkent city hosted 99,151 active enterprises, against a median of 27,634 for the remaining thirteen territories, whereas the agrarian and densely populated regions of the Fergana Valley and the south combined moderate enterprise stocks with high informality. Whether this pattern reflects a systematic, quantifiable inverse relationship between business density and informality, or is instead produced by the scale of regional output and a few structurally distinctive territories, has not been established with regional data for Uzbekistan.

The aim of this study is to measure, with official regional statistics, the elasticity of the shadow economy share with respect to the stock of active enterprises and the scale of gross regional product across the 14 administrative territories of Uzbekistan. The specific objectives are: (i) to compile and harmonize a territorial dataset from four official series; (ii) to describe and contrast high-formalization hubs with high-informality territories; (iii) to estimate a multi-variable log-linear ordinary least squares (OLS) model and evaluate it against standard diagnostic tests; (iv) to assess the robustness of the estimates to influential observations, panel structure and specification; and (v) to derive implications for regional policy. The object of the research is the regional shadow economy of Uzbekistan; the subject is its statistical relationship with active business density and regional output.

Two hypotheses were formulated before estimation. H1: a higher stock of active enterprises suppresses the shadow economy share ($\beta 1 < 0$). H2: a larger gross regional output contributes to institutional formalization and therefore lowers the shadow economy share ($\beta 2 < 0$). The scientific novelty lies in three elements. First, four official territorial series are integrated into one reproducible dataset, and elasticities rather than simple correlations are estimated for Uzbekistan's territories. Second, the count of registered enterprises is separated from the scale of the regional economy, two dimensions that are often treated jointly as a single business-environment factor. Third, the cross-sectional estimates are subjected to influential-observation, panel fixed-effects and long-difference checks that are rarely reported in territorial studies based on very small samples.

LITERATURE REVIEW

The modern empirical literature was consolidated by the survey of Schneider and Enste, which linked the size of the shadow economy to tax and social-security burdens, regulation and labor-market conditions and catalogued the direct, indirect and model-based methods used to measure it [1]. Later reviews showed that the widely used multiple indicators multiple causes (MIMIC) approach, applied by Medina and Schneider to a large panel of countries, produces estimates that depend on the chosen causes and indicators and on calibration to an external benchmark [2, 5]. Gërxhani emphasized that definitions and dominant mechanisms differ between developed and less developed countries, so that findings should be compared with care [8]. For transition economies, Kaufmann and Kaliberda documented that unofficial activity absorbed a substantial part of the output that official statistics failed to record during the post-socialist recession [9]. A critical implication is that regional figures produced by national-accounts procedures, as in Uzbekistan, are conceptually different from MIMIC

indices and should not be benchmarked directly against them.

Two further considerations are relevant for the Uzbek case. Official regional figures rely on the national-accounts treatment of the non-observed economy, which combines household and enterprise surveys with administrative records, and they express the shadow economy as a share of gross regional product rather than as a latent index. Such estimates are arguably less exposed to the calibration problem of MIMIC models, but they are more exposed to dependence on the very registers, such as the enterprise register, that a researcher may wish to use as explanatory variables. This tension is seldom discussed in the literature reviewed here and is taken up in the methodology and the discussion.

Competing schools explain why producers remain unregistered. Dell'Anno's survey distinguishes dualist, structuralist, legalist and voluntarist perspectives [10]. The legalist view associated with De Soto attributes informality to the high monetary and time cost of complying with bureaucratic procedures, so that lowering entry barriers should draw producers into the formal sector [11]. Cross-country evidence supports part of this reasoning: heavier entry regulation is associated with more corruption and larger unofficial economies without improvements in product quality [12], while regulatory discretion and corruption appear to matter at least as much as nominal tax rates [7]. Loayza's model similarly ties informality to the interaction between tax burdens, enforcement and the quality of government services [6]. The dualist position of La Porta and Shleifer is more skeptical: informal firms are small, managed by owners with little education and far less productive than formal firms, so lighter regulation alone will not turn them into formal enterprises, and informality declines mainly as economies develop [3, 4]. Experimental and quasi-experimental evidence reviewed by Bruhn and McKenzie is consistent with this caution, since simplified registration typically generates only modest increases in formalization [13].

The two views generate contrasting empirical expectations. If the legalist mechanism dominates, territories with more enterprises per unit of output should be those in which formalization has been easiest and informality is lowest. If the dualist mechanism dominates, informality should be governed by the scale, productivity and skill content of the local economy, and the number of registered units should add little once these are controlled. Existing studies rarely confront these expectations with the same data, and the present design allows them to be compared directly.

The expected sign of the relationship between the number of registered enterprises and informality is less obvious than is often assumed. On the one hand, a dense formal business population widens the tax base, raises the probability of enforcement, offers formal supply-chain linkages and intensifies competition against unregistered rivals, which is the mechanism envisaged by H1. Klapper, Laeven and Rajan showed that entry regulation reduces the creation of new firms in sectors where entry would otherwise be frequent and shapes the size and growth of incumbent firms [14]. On the other hand, Ulyssea demonstrated that informality operates at two margins: unregistered firms, and registered firms that employ workers informally or under-report activity, so that a higher count of registered firms can coexist with substantial informal activity, and lowering entry costs has limited effects on aggregate informality [15, 16]. Because registered enterprises in developing economies include many micro-units, the quality and scale of the enterprise base may matter more for formalization than its numerical size [3, 16].

Sectoral composition complicates the picture further. Territories with large agricultural and household-based service sectors typically contain many small producers who are hard to observe and register, whereas territories dominated by capital-intensive extraction or heavy industry are characterized by a small number of large, easily monitored taxpayers. The same enterprise count may therefore correspond to very different degrees of formal-sector depth [4, 16]. A cautious reading of the literature is that the count of registered firms is neither necessary nor sufficient for a low shadow share, and that its sign in a regression depends on what else is held constant, which motivates the multi-variable specification adopted here.

Four limitations of the existing literature motivate the present study. First, most evidence is cross-national or national; regional heterogeneity within a single transition economy is rarely

examined, and studies of the post-socialist space have concentrated on national aggregates [8, 9]. Second, official regional estimates of the informal and hidden economy in Uzbekistan were first published only in August 2024, which explains the absence of territorial econometric evidence. Third, empirical designs seldom separate the number of formal enterprises from the scale of the local economy, or test whether an apparent business-density effect survives once a handful of extreme observations, such as a capital city or a resource-based region, are set aside. Fourth, the interaction between registry-based explanatory variables and shadow estimates built from surveys and administrative data has not been examined, so it remains unclear how far a statistical association between the two reflects behavior and how far it reflects measurement. The present paper addresses these limitations with an explicit log-log specification, diagnostic testing and robustness analysis for the 14 territories of Uzbekistan, and it reports results as elasticities that can be compared directly with future estimates for other transition economies.

METHODOLOGY

A quantitative, observational design was adopted. Four official datasets published by the National Statistics Committee of Uzbekistan (stat.uz) were downloaded in spreadsheet form: (1) the annual number of active enterprises and organizations by territory; (2) the share of the informal and hidden economy in GDP (GRP) by territory, in percent; (3) the volume of the informal and hidden economy by territory, in billion soum; and (4) gross regional product at current prices, in billion soum. The territorial unit of analysis comprised the Republic of Karakalpakstan, 12 regions (viloyat) and Tashkent city, that is, 14 units identified by their four-digit classifier codes. District-level and city-level rows contained in the enterprise file were excluded, and the national row was used only for context. The datasets were merged on classifier code and year.

The shadow-economy series begin in 2017; entries of zero for 2010–2016 in the source files denote unpublished values and were treated as missing. The enterprise series extend to 2026, but only 2017–2025 overlapped with the other series, giving a balanced panel of 14 × 9 = 126 territory-years. The year 2024 was fixed in advance as the reference cross-section, because it is the most recent year before the marked decline in the national shadow share (from 34.3% to 26.9%) and in the enterprise stock that appears in 2025 and that the metadata do not attribute to a specific methodological change. The 2024 GRP values are flagged by the publisher as preliminary. The accounting consistency of the files was verified by recomputing each territory's share as volume divided by GRP, which reproduced the published share to within 0.05 percentage points in every case.

Model specification and variables. The following multi-variable log-linear model was estimated for the cross-section of territories i:

$$ln(Shadow\_Share_i) = \beta_0 + \beta_1\, ln(Enterprises_i) + \beta_2\, ln(GRP_i) + \varepsilon_i$$

Shadow_Share was the share of the informal and hidden economy in GRP, in percent; Enterprises was the number of active enterprises and organizations; and GRP was gross regional product at current prices in billion soum. The log-log form was chosen for three reasons: the coefficients are elasticities that are independent of measurement units; the logarithmic transformation reduces the pronounced right-skewness of the enterprise and GRP variables (GRP in Tashkent city was 10.6 times that of Syrdarya region); and it allows the effect of a percentage change in business density to be stated directly. Under H1 and H2 both β1 and β2 were expected to be negative. Because the dependent variable equals shadow volume divided by GRP, an additional specification with the logarithm of shadow volume as the dependent variable was estimated; algebraically it must return the same enterprise coefficient and a GRP coefficient larger by exactly one, which served as a computational check.

Estimation and diagnostic procedures. Parameters were estimated by OLS with conventional standard errors, and heteroscedasticity-consistent (HC3) standard errors were reported alongside because the sample contained only 14 observations and 11 residual degrees of freedom. Multicollinearity was assessed with variance inflation factors (VIF), using 5 as a conservative and 10 as a conventional threshold. Heteroscedasticity was tested with the Breusch–Pagan [17] and White

[18] tests, normality of residuals with the Jarque–Bera test [19] and the Shapiro–Wilk test, and functional form with Ramsey's RESET test using squared fitted values. Influential observations were identified from Cook's distance (threshold 4/n), leverage (threshold 2k/n, where k = 3 estimated parameters) and externally studentized residuals (absolute value above 2).

Robustness was examined in four ways. First, the model was re-estimated after removing Navoi, Tashkent city, and both. Second, because the two elasticities were expected to be similar in magnitude and opposite in sign, a Wald test of H0: $\beta 1 + \beta 2 = 0$ was conducted, and a restricted model with the logarithm of GRP per enterprise as the single regressor was estimated. Third, using the 2017–2025 panel, pooled OLS, pooled OLS with year dummies and two-way (territory and year) fixed-effects models were estimated with standard errors clustered by territory [20]; the fixed-effects models were repeated for 2017–2024 to exclude the 2025 discontinuity, and year-by-year cross-sections were estimated. Fourth, a long-difference model relating the change in ln(Shadow_Share) between 2017 and 2024 to the changes in ln(Enterprises) and ln(GRP) was estimated. Since 14 clusters are few, the clustered standard errors were treated as indicative. All computations were performed in Python 3 (pandas, statsmodels, SciPy), and the merged dataset is reproduced in Table 2, so that every estimate can be replicated. The significance level was 5%.

ANALYSIS AND RESULTS

Descriptive statistics and regional contrasts. Table 1 summarizes the 14 territories in 2024 and the pooled panel. The mean territorial shadow share was 44.6% (standard deviation 11.4 percentage points; median 49.2%), with a minimum in Navoi and a maximum in Surkhandarya. The unweighted mean shadow share across territories declined from 56.2% in 2017 to 44.6% in 2024.

Table 1.

Descriptive statistics, 14 territories, 2024 (panel rows: 2017–2025)

Variable	Mean	Std. Dev.	Min	Max
Shadow share of GRP, %	44.64	11.41	17.5	56.2
Active enterprises, number	34,645	20,691	13,432	99,151
GRP, bn soum	95,216.6	68,752.4	28,655.5	302,878.5
Shadow economy volume, bn soum	37,666.2	16,666.4	11,424.0	76,421.6
GRP per enterprise, bn soum	2.71	1.04	1.79	5.96
Panel 2017–2025 (N = 126): shadow share, %	46.65	11.91	13.1	65.8
Panel: active enterprises	30,016	19,729	8,771	117,812
Panel: GRP, bn soum	59,956.7	52,540.9	8,115.8	367,222.3

The contrast between hubs and high-informality territories is shown in Table 2. Tashkent city recorded 99,151 active enterprises, GRP of 302,878.5 billion soum and a shadow share of 25.2%. Navoi recorded 21,880 enterprises, GRP of 130,439.5 billion soum and the lowest shadow share, 17.5%. The three highest-informality territories, Surkhandarya (56.2%), Khorezm (54.9%) and Namangan (53.0%), had 24,657, 27,089 and 27,634 enterprises and GRP of 56,553.9, 53,640.0 and 71,869.1 billion soum. The simple mean of the two hubs was 21.4% for the shadow share, 60,516 for enterprises and 4.51 billion soum for GRP per enterprise; the corresponding means for the three high-informality territories were 54.7%, 26,460 and 2.29 billion soum. Navoi's GRP per enterprise (5.96 billion soum) was the highest and Karakalpakstan's (1.79) the lowest in the sample.

Table 2.

Regional data used in the estimation, 2024 (sorted by shadow share)

Territory	Shadow share, %	Active enterprises	GRP, bn soum	Shadow volume, bn soum	GRP per enterprise, bn soum
Surkhandarya	56.2	24,657	56,553.9	31,782.1	2.29
Khorezm	54.9	27,089	53,640.0	29,439.6	1.98

Namangan	53.0	27,634	71,869.1	38,076.1	2.60
Kashkadarya	52.9	33,064	83,950.0	44,451.2	2.54
Jizzakh	49.6	19,923	43,545.9	21,596.8	2.19
Samarkand	49.5	43,477	103,882.7	51,443.5	2.39
Fergana	49.2	42,574	93,605.7	46,010.1	2.20
Bukhara	49.1	31,203	73,434.1	36,078.6	2.35
Karakalpakstan	47.0	25,928	46,299.8	21,774.6	1.79
Andijan	45.5	29,717	93,026.3	42,355.3	3.13
Syrdarya	39.9	13,432	28,655.5	11,424.0	2.13
Tashkent region	35.5	45,295	151,251.8	53,672.3	3.34
Tashkent city	25.2	99,151	302,878.5	76,421.6	3.05
Navoi	17.5	21,880	130,439.5	22,801.5	5.96

All 14 territories recorded lower shadow shares in 2024 than in 2017. The largest declines were in Navoi (37.2% to 17.5%, −19.7 percentage points), Andijan (−19.0 points) and Syrdarya (−15.6 points), and the smallest in Karakalpakstan (−4.6 points), Kashkadarya (−7.0) and Khorezm (−7.5). Over the same period the number of active enterprises grew by 80% in Tashkent city, 149% in Navoi, 47% in Namangan and 13% in Andijan, and GRP at current prices grew by a factor of 5.2 in Tashkent city, 7.1 in Navoi and 3.9 in Namangan.

In logarithms, the Pearson correlation of ln(Shadow_Share) with ln(Enterprises) was −0.204, with ln(GRP) −0.578, and the correlation between ln(Enterprises) and ln(GRP) was 0.867. Spearman rank correlations of the shadow share were −0.160 (p = 0.584) with enterprises, −0.490 (p = 0.075) with GRP and −0.459 (p = 0.098) with GRP per enterprise. Figure 1 plots the shadow share against enterprises and against GRP per enterprise.

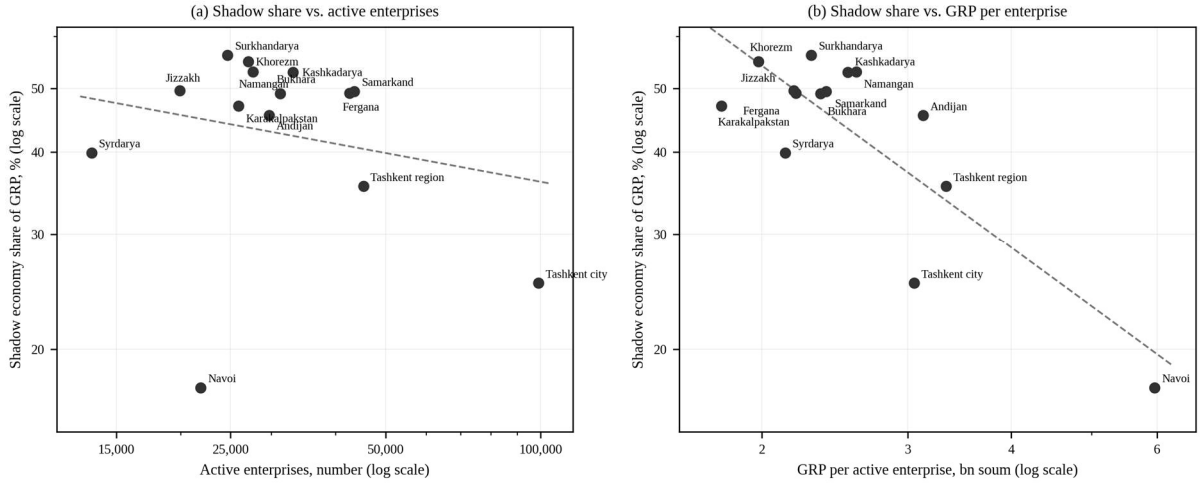

Figure 1. Shadow economy share, active enterprises and GRP per enterprise, 14 territories, 2024 (dashed lines: log-log OLS fits)

Table 3 reports the baseline estimates. The model was jointly significant, with $R^2 = 0.688$ and adjusted $R^2 = 0.632$.

Table 3.

OLS estimates of the baseline model, dependent variable ln(Shadow_Share), 2024

Variable	Coeff.	Std. error	t-stat.	p-value	95% CI	HC3 s.e.	HC3 p
Constant	5.210	1.239	4.20	0.001	[2.483; 7.937]	3.330	0.146
ln(Enterprises)	0.850	0.240	3.54	0.005	[0.321; 1.380]	0.738	0.274
ln(GRP)	−0.908	0.190	−4.78	<0.001	[−1.326; −0.489]	0.558	0.132

The coefficient on ln(GRP) was −0.908 (standard error 0.190, t = −4.78, p = <0.001), and the coefficient on ln(Enterprises) was 0.850 (standard error 0.240, t = 3.54, p = 0.005). In percentage terms, a 1% increase in GRP was associated with a change of −0.899% in the shadow share, and a 1% increase in active enterprises with a change of +0.850%, holding the other regressor constant. With HC3 standard errors neither slope was significant at the 5% level (p = 0.274 and 0.132).

In the bivariate regressions, the coefficient on ln(Enterprises) was −0.145 (standard error 0.201, p = 0.484; R^2 = 0.042), corresponding to a 0.144% decline in the shadow share per 1% increase in enterprises, while the coefficient on ln(GRP) was −0.325 (standard error 0.133, p = 0.031; R^2 = 0.334). The Wald test of β1 + β2 = 0 in the baseline model returned F(1, 11) = 0.224 (p = 0.645). The restricted model with ln(GRP per enterprise) as the only regressor produced a slope of −0.921 (standard error 0.182, p < 0.001), R^2 = 0.682, adjusted R^2 = 0.655 and F(1, 12) = 25.72. In the volume specification, the coefficient on ln(Enterprises) was 0.852 (p = 0.005) and that on ln(GRP) was 0.091 (p = 0.640), with R^2 = 0.852; the coefficients differ from the baseline by 0.002 and by 0.999, respectively, the small residual deviations reflecting rounding of the published shares.

Table 4.

Diagnostic tests for the baseline model, 2024

Test	Statistic	p-value	Outcome at 5%
Variance inflation factor (both regressors)	4.03	—	< 5
Breusch–Pagan LM	0.114	0.945	No heteroscedasticity rejected
White test LM	7.013	0.220	No heteroscedasticity rejected
Jarque–Bera	1.150	0.563	Normality not rejected
Shapiro–Wilk W	0.914	0.181	Normality not rejected
Ramsey RESET (squared fitted values), F(1, 10)	7.817	0.019	Linearity rejected at 5%

The VIF was 4.03 for both regressors. Neither the Breusch–Pagan test (LM = 0.114, p = 0.945) nor the White test (LM = 7.013, p = 0.220) rejected homoscedasticity, and neither the Jarque–Bera test (p = 0.563) nor the Shapiro–Wilk test (p = 0.181) rejected normality of the residuals. The RESET test rejected the linear log-log form (F = 7.817, p = 0.019). Cook's distance exceeded the threshold of 0.286 for Navoi (3.198), Tashkent city (2.155) and Syrdarya (0.493). Leverage exceeded 2k/n = 0.429 for Navoi (0.782) and Tashkent city (0.552). Only Tashkent city had an absolute externally studentized residual above 2 (−3.02); the value for Syrdarya was −1.97.

Table 5.

Elasticities of the shadow share under alternative specifications

Specification	β_1: ln(Enterprises)	β_2: ln(GRP)	N	R^2
(1) Bivariate: enterprises only	−0.145 (0.201)	—	14	0.042
(2) Bivariate: GRP only	—	−0.325 (0.133)**	14	0.334
(3) Baseline, 2024	0.850 (0.240)***	−0.908 (0.190)***	14	0.688
(4) Excluding Navoi	0.168 (0.440)	−0.363 (0.350)	13	0.414
(5) Excluding Tashkent city	1.097 (0.200)***	−0.874 (0.145)***	13	0.794
(6) Excluding both	0.480 (0.322)	−0.395 (0.246)	12	0.223
(7) Pooled OLS, 2017–2025	0.354 (0.267)	−0.497 (0.168)***	126	0.511
(8) Pooled with year dummies	0.605 (0.267)**	−0.774 (0.219)***	126	0.618

(9) Two-way fixed effects	0.019 (0.106)	−0.698 (0.143)***	126	0.989
(10) Two-way FE, 2017–2024	−0.019 (0.121)	−0.683 (0.134)***	112	0.988
(11) Long difference, 2017–2024	0.073 (0.127)	−0.703 (0.117)***	14	0.770

Table 5 summarizes the alternative specifications. Excluding Navoi reduced the enterprise coefficient to 0.168 ($p = 0.711$) and the GRP coefficient to −0.363 ($p = 0.324$), with adjusted $R^2 = 0.297$. Excluding Tashkent city left the coefficients at 1.097 ($p < 0.001$) and −0.874 ($p < 0.001$), with adjusted $R^2 = 0.752$. Excluding both territories gave insignificant coefficients of 0.480 and −0.395. The GRP coefficient was negative in all ten specifications that included it. In the panel, it was significant in the pooled model ($p = 0.003$), the year-dummy model ($p < 0.001$) and both fixed-effects models ($p < 0.001$), whereas the enterprise coefficient was insignificant in the pooled and fixed-effects models and positive and significant ($p = 0.023$) only in the pooled model with year dummies and no territory effects.

Year-by-year cross-sections showed that the enterprise coefficient rose from 0.209 in 2017 ($p = 0.369$) to 0.872 in 2022 ($p = 0.009$) and 0.807 in 2025 ($p = 0.002$), while the GRP coefficient moved from −0.371 in 2017 ($p = 0.120$) to −0.984 in 2025 ($p = <0.001$), and cross-sectional R^2 increased from 0.273 to 0.762. In the long-difference model, the GRP coefficient was significant ($p < 0.001$), the enterprise coefficient was not ($p = 0.576$), and $F(2, 11) = 18.41$ ($p < 0.001$).

DISCUSSION

The results answer the two hypotheses differently. H1, which predicted that a larger stock of active enterprises would suppress the shadow share, was not supported. The bivariate elasticity was small, negative and statistically indistinguishable from zero, and the conditional elasticity was positive. H2 received qualified support: the scale of regional output was negatively related to the shadow share in every specification and significantly so in the pooled, fixed-effects and long-difference estimates. This pattern is consistent with the dualist reading of the international literature, in which informality is a feature of low-productivity economic structures and declines mainly with development rather than with the multiplication of registered units [3, 4].

The positive enterprise coefficient in the baseline model must be read together with the very high correlation between enterprises and GRP (0.867 in logarithms), and with the near equality of the two elasticities in absolute value. The Wald test could not reject $\beta 1 + \beta 2 = 0$, and the restricted model with ln(GRP per enterprise) explained a similar share of the variance with a single regressor. Substantively, the shadow share was lower where the output generated per registered enterprise was larger. This is compatible with evidence that formal firms differ widely in size and productivity and that informal producers coexist with formal ones [15, 16]. A large count of registered entities, many of which in developing economies are micro-units, is therefore a weak proxy for the depth of the formal sector [3]. That the enterprise coefficient was near zero in the fixed-effects and long-difference models, where between-territory differences are removed, and that it disappeared without Navoi, indicates that the conditional positive sign should not be interpreted as a causal effect of enterprise proliferation on informality.

The volume specification adds an important qualification to H2. Shadow volume was strongly and nearly proportionally related to the number of enterprises (elasticity 0.852) but essentially unrelated to GRP once enterprises were controlled (0.091, $p = 0.640$). Because the shadow share is volume divided by GRP, the negative GRP elasticity of the share arises largely because GRP differs far more across territories than does shadow volume. It therefore partly reflects the accounting structure of the indicator and should not be attributed entirely to institutional formalization. The methodological notes of the publisher state that the underlying estimates rely on surveys and administrative data, which may create a further dependence between the estimated shadow volume

and the registered enterprise base.

The territorial contrasts fit this interpretation. Tashkent city plausibly combines the densest concentration of registered enterprises, the head offices and banking infrastructure of large formal firms, and the most developed tax and administrative institutions, so that more economic activity passes through channels in which concealment is costly. Navoi's low shadow share coincides with the highest GRP per enterprise in the sample, which is consistent with an output structure dominated by a small number of large formal producers in heavy industry and mining, where the scope for informal production is limited. In Surkhandarya, Khorezm and Namangan, by contrast, GRP per enterprise was among the lowest, enterprise numbers were moderate, and a large agrarian and household-based labor supply is served by fragmented trade and service activity. Under the legalist reasoning of De Soto and the entry-cost evidence of Djankov and coauthors, producers in such settings weigh registration costs against modest expected gains [11, 12], and Friedman and coauthors suggest that discretion and corruption in enforcement may reinforce that calculation [7].

The panel evidence points in the same direction. Between 2017 and 2024 Navoi's GRP grew by about 1.96 log points and its shadow share fell by 0.75 log points, whereas Karakalpakstan's GRP grew by 1.21 log points and its share fell by only 0.09. The long-difference elasticity of −0.703 summarizes this association: within territories, faster growth of output coincided with steeper formalization, while faster growth of the enterprise stock did not add explanatory power. Compared with the international literature, this suggests that the Uzbek regional pattern resembles the development-driven decline of informality described by La Porta and Shleifer more closely than the regulation-driven reallocation envisaged in the legalist tradition, although the reforms of the period may have contributed to both output growth and formalization and cannot be separated here [3, 11].

Several policy levers follow, with the caveat that the estimates are associations. First, digitalization of tax compliance, including electronic invoicing, online reporting and data-matching between registries, raises the detection probability and lowers compliance costs simultaneously; this fits the emphasis on enforcement quality found in the models of Loayza and Ulyssea [6, 15]. Second, SME incubation should aim at scaling and productivity, not only at registering new units, since output per enterprise rather than enterprise numbers was the statistically relevant margin. Third, reducing compliance transaction costs, through one-stop registration and licensing and predictable inspections, is warranted, but the modest effects of registration reforms reported by Bruhn and McKenzie suggest that such measures should be combined with incentives and enforcement [13]. Fourth, regional programmes should be differentiated: hubs such as Tashkent city require measures against the hidden component of formal firms' activity, whereas valley and southern territories require productivity-enhancing investment and non-farm employment.

The study has limitations. The cross-section contains only 14 observations, so inference is sensitive to individual territories, as the influence diagnostics and the exclusion of Navoi demonstrate; the RESET test also indicated that a simple log-linear form is an approximation. The enterprise count was not normalized by population or by the number of working-age residents, which are not included in the datasets used. GRP is measured at current prices, so that inflation contributes to both trends, although year effects absorb common price movements in the panel. The 2025 figures show a discontinuity that the metadata do not explain, and the 2024 GRP figures are preliminary. Spatial dependence between neighboring territories, sectoral composition and other determinants such as tax burden, corruption and education were not modeled. Finally, only 14 clusters were available for clustered inference. These limitations preclude causal claims and indicate that the estimates describe stable statistical regularities rather than policy effects.

CONCLUSION

This study assessed the relationship between active business density, regional output and the shadow economy across the 14 territories of Uzbekistan using four official statistical series and a log-log OLS framework. The principal finding is that the shadow share is lower where regional output is larger, and more precisely where output per registered enterprise is larger, whereas the number of

registered enterprises does not, by itself, reduce informality. In the 2024 cross-section the baseline model explained 68.8% of the variance, with an output elasticity of −0.908 and an enterprise elasticity of +0.850; in bivariate form, a 1% increase in enterprises was associated with only a 0.14% decline in the shadow share, which was not statistically distinguishable from zero.

With respect to the hypotheses, H1 was not supported: no robust inverse relationship between the count of active enterprises and the shadow share was found, and the conditional positive coefficient disappeared in fixed-effects and long-difference models and without Navoi. H2 was supported with reservations: the negative association with GRP held in sign across all specifications and in significance across the panel and long-difference estimates, although it is partly an arithmetic consequence of measuring the shadow economy as a share of GRP and is sensitive to Navoi in the cross-section. The empirical contribution is to show that, for Uzbekistan's territories, the relevant margin of formalization is the scale and productivity of the formal enterprise base rather than its numerical size.

For regional economic planning, the findings recommend that formalization policy move beyond counting new registrations. Digital tax compliance systems, SME support directed at scaling and productivity, lower transaction costs of compliance and regionally differentiated measures for the valley and southern territories should be prioritized, with monitoring based on output per enterprise and on shadow volume in addition to shadow share. Future research should use district-level data, which would greatly enlarge the sample; normalize enterprise numbers by population and sector; apply spatial econometric models and instrumental strategies to address dependence and endogeneity; compare official estimates with MIMIC-based indices; and extend the panel as further years of consistently defined data become available.

A further practical implication concerns territories such as Namangan region, where the 2024 shadow share of 53.0% coincided with GRP per enterprise of 2.60 billion soum, below the sample mean of 2.71. For such territories, the results suggest that raising the productivity and scale of registered enterprises, and thereby the formal output generated per unit, is a more promising route to reducing informality than expanding the number of registrations alone, while acknowledging that the estimates describe associations that require confirmation with richer data.

REFERENCES

1. Schneider F., Enste D.H. Shadow economies: size, causes, and consequences // Journal of Economic Literature. — 2000. — Vol. 38, No. 1. — P. 77–114.
2. Medina L., Schneider F. Shadow economies around the world: what did we learn over the last 20 years? — Washington, DC: International Monetary Fund, 2018. — 76 p. — (IMF Working Paper; No. 18/17).
3. La Porta R., Shleifer A. Informality and development // Journal of Economic Perspectives. — 2014. — Vol. 28, No. 3. — P. 109–126.
4. La Porta R., Shleifer A. The unofficial economy and economic development // Brookings Papers on Economic Activity. — 2008. — No. 2. — P. 275–352.
5. Schneider F., Buehn A. Shadow economy: estimation methods, problems, results and open questions // Open Economics. — 2018. — Vol. 1, No. 1. — P. 1–29.
6. Loayza N.V. The economics of the informal sector: a simple model and some empirical evidence from Latin America // Carnegie-Rochester Conference Series on Public Policy. — 1996. — Vol. 45. — P. 129–162.
7. Friedman E., Johnson S., Kaufmann D., Zoido-Lobaton P. Dodging the grabbing hand: the determinants of unofficial activity in 69 countries // Journal of Public Economics. — 2000. — Vol. 76, No. 3. — P. 459–493.
8. Gërxhani K. The informal sector in developed and less developed countries: a literature survey // Public Choice. — 2004. — Vol. 120, No. 3–4. — P. 267–300.
9. Kaufmann D., Kaliberda A. Integrating the unofficial economy into the dynamics of post-socialist economies: a framework of analysis and evidence // Economic Transition in Russia and the New

States of Eurasia / ed. by B. Kaminski. — Armonk, NY: M.E. Sharpe, 1996. — P. 81–120.
10. Dell'Anno R. Theories and definitions of the informal economy: a survey // Journal of Economic Surveys. — 2022. — Vol. 36, No. 5. — P. 1610–1643.
11. De Soto H. The Other Path: The Invisible Revolution in the Third World. — New York: Harper & Row, 1989. — 271 p.
12. Djankov S., La Porta R., Lopez-de-Silanes F., Shleifer A. The regulation of entry // Quarterly Journal of Economics. — 2002. — Vol. 117, No. 1. — P. 1–37.
13. Bruhn M., McKenzie D. Entry regulation and the formalization of microenterprises in developing countries // World Bank Research Observer. — 2014. — Vol. 29, No. 2. — P. 186–201.
14. Klapper L., Laeven L., Rajan R. Entry regulation as a barrier to entrepreneurship // Journal of Financial Economics. — 2006. — Vol. 82, No. 3. — P. 591–629.
15. Ulyssea G. Firms, informality, and development: theory and evidence from Brazil // American Economic Review. — 2018. — Vol. 108, No. 8. — P. 2015–2047.
16. Ulyssea G. Informality: causes and consequences for development // Annual Review of Economics. — 2020. — Vol. 12. — P. 525–546.
17. Breusch T.S., Pagan A.R. A simple test for heteroscedasticity and random coefficient variation // Econometrica. — 1979. — Vol. 47, No. 5. — P. 1287–1294.
18. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity // Econometrica. — 1980. — Vol. 48, No. 4. — P. 817–838.
19. Jarque C.M., Bera A.K. A test for normality of observations and regression residuals // International Statistical Review. — 1987. — Vol. 55, No. 2. — P. 163–172.
20. Wooldridge J.M. Econometric Analysis of Cross Section and Panel Data. — 2nd ed. — Cambridge, MA: MIT Press, 2010. — 1064 p.