Prediction Versus Identification: Machine Learning and Small-Sample Inference for Public Credit Guarantees to Small and Medium-Sized Enterprises in Seven European Countries

Abdellah HAFID[1], Anouar RIAD SOLH[2], Hamid EZ-ZAHRAOUY[3]
Mohammed V University of Rabat, Faculty of Sciences, Morocco[1, 2, 3]
Laboratory of Condensed Matter and Interdisciplinary Sciences (LaMCScI-URL-CNRST)-Faculty of Sciences, Mohammed V University, Rabat, Morocco[1, 3]
Laboratory of Conception and Systems (Electronics, Signals and Informatics)-Faculty of Sciences, Mohammed V University, Rabat, Morocco[2]

***Abstract*—Public credit guarantees for small and medium-sized enterprise (SME) lending, long-standing instruments of financing policy, have seen their use intensify across Europe during episodes of credit stress, particularly during the pandemic. Their effect on credit volumes nonetheless remains poorly established at the aggregate level. This article examines whether these schemes genuinely stimulate lending or play a primarily stabilizing role, using a panel of seven European countries observed from 2010 to 2024. The contribution is twofold and methodological. First, on a real policy-evaluation case, resampling-based inference adapted to a small number of clusters is shown to overturn the conclusions that standard inference would support. Second, the analysis separates prediction from causal identification: a debiased double machine learning estimator and a SHAP analysis illustrate that the nonparametric flexibility of machine learning does not create identifying power where it is structurally absent, and that predictive salience must not be conflated with causal significance. SME credit is persistent and is the parameter closest to being identifiable under valid inference, though it falls short of significance at the 5 per cent level. The guarantee ratio is not detectable once inference is corrected, and no coefficient is robustly identified at the 5 per cent level. A size-calibrated power analysis shows that the design could not reliably detect effects smaller than about 13 per cent per standard deviation, more than twice the estimated effect, so non-detectability reflects a limit of statistical power at seven clusters rather than evidence of no effect. The overall pattern is consistent with an endogenous response of guarantees to credit conditions, and does not establish additionality.**

***Keywords*—SME credit; public credit guarantees; wild cluster bootstrap; resampling-based inference; statistical power; double machine learning; prediction versus identification; dynamic panel**

I. Introduction

The expansion of public credit guarantee schemes for SME lending — instruments whose use intensified during episodes of credit stress, particularly during the pandemic — has been one of the most consistent economic-policy choices of the past decade in Europe. Yet a financial-stability question remains open: do these schemes genuinely stimulate new lending, or do they play a primarily stabilizing role in times of stress? The OECD [1] documents a rise in the median share of state-guaranteed loans in the SME loan stock, from 6.9 percent in 2021 to 8.0 percent in 2022, following a marked acceleration during the pandemic. This trend points to a central hypothesis, still little explored on recent aggregate data: guarantees are presumed to stimulate new lending beyond what would have occurred without them — additionality. The World Bank (Carvajal and Didier [2]) and the OECD [1] document the growth of guarantee schemes and propose design principles as well as methodological recommendations to improve their monitoring. These reference works open the way to aggregate empirical analyses of the post-pandemic phases, still scarce in the literature; it is to this space that the present article seeks to contribute. Rather than claiming to settle the additionality question — which a seven-country panel cannot do — this policy question is taken as a testing ground for a methodological demonstration: showing what inference adapted to a small number of clusters changes in the conclusions, in a case where the policy stakes are real.

The academic literature on this point remains fragmented. Existing evaluations are most often national and cover pre-pandemic periods (Lelarge, Sraer and Thesmar [3]; Boschi et al. [4]). Aggregate studies on a multi-period European panel that incorporate the post-2010 sequence remain rare. On the methodological side, a difficulty rarely addressed concerns statistical inference in a low cross-sectional dimension: standard econometric estimates produce p-values whose asymptotic validity assumes large N, an assumption scarcely plausible with seven countries. Cameron, Gelbach and Miller [5] and Webb [6] showed that the wild cluster bootstrap (WCB) is the recommended inference procedure in such configurations; yet its systematic application remains rare in the empirical literature on SME finance. This inference question is still little discussed in the field, which calls for cautious interpretation of published results on small panels. Documenting this question on a real case, rather than by simulation, is the primary contribution of this article: the non-detectability of the effects reported here is not a measurement

failure but the substantive result produced by correctly calibrated inference.

This article contributes to both debates and makes two distinct methodological contributions. On substance, it documents that the persistence of SME credit — $\rho = 0.629$, with a half-life of 1.50 years — is the parameter closest to identifiable under valid inference with seven clusters ($p_{WCB} = 0.065$, short of significance at the 5% level); once inference is corrected, neither the guarantee ratio nor institutional governance is robustly identified (only persistence is nominally significant in a first-difference specification, $p_{WCB} = 0.026$; governance remains non-significant even in first differences; Table XXIII, Appendix N), although their signs are consistent with the hypotheses. The first methodological contribution is the systematic confrontation of HC1 and WCB p-values, which reveals striking gaps — from 0.018 to 0.15 on governance and from <0.001 to 0.065 on persistence — and concretely documents the pitfall of naive inference at low N. The second belongs to machine learning applied to causal inference: following the distinction between predicting and identifying [7, 8], the debiased double machine learning of Chernozhukov et al. [9] and a SHAP analysis [10] are deployed to show, on real data, that nonparametric functional flexibility does not create cluster-valid identifying power where it is structurally absent at N = 7, and that a variable's predictive importance must not be conflated with its causal significance. These two contributions are complementary: the empirical result on the non-identifiability of guarantees is more credible precisely because it is produced with correct inference and probed with orthogonalized machine-learning methods.

These two dimensions — factual and methodological — structure the empirical choices described from Section III onward.

II. Literature Review

A. Information Asymmetries and Credit Guarantees

The theoretical justification for public guarantees rests on the seminal work of Stiglitz and Weiss [11] on credit rationing under information asymmetry. In that framework, some creditworthy firms are denied loans that fully informed banks would grant them. Public guarantees mitigate this imperfection by transferring default risk to a public third party, giving banks an incentive to lend more. Bester [12] and Besanko and Thakor [13] specify the conditions under which guarantees separate good borrowers from bad ones, reducing adverse selection.

Berger and Udell [14] and Boot [15] show that SME lending relies on the gradual accumulation of proprietary information within the banking relationship, creating strong temporal dependence in credit volumes. Petersen and Rajan [16] confirm that long-term banking relationships give SMEs access to more favorable terms — a relational persistence that compounds the volume persistence confirmed by the data ($\rho = 0.629$). To the authors' knowledge, no work has empirically examined how this structural persistence affects the statistical detectability of guarantee additionality in a low-N aggregate panel. The results suggest precisely that this informational inertia constitutes a major methodological obstacle: the effect of guarantees, if it exists, diffuses through an already strongly autoregressive credit system, which renders it indistinguishable from zero under valid inference with seven clusters.

B. Empirical Evaluations of Additionality

Empirical evaluations of guarantee additionality reach contrasting conclusions (Cusmano [17]). The literature review of Gozzi and Schmukler [18] concludes that additionality is positive but modest at the microeconomic level, with strong heterogeneity depending on the institutional framework and the level of guarantee coverage. On French firm-level data, Lelarge, Sraer and Thesmar [3] identify a positive effect on the growth of beneficiary firms, but a simultaneous rise in default risk, suggesting a compositional transfer rather than net additionality. For Italy, Boschi et al. [4] document real but modest volume additionality, and show that the guarantee's coverage level is the key determinant of its effectiveness. Asdrubali and Signore [19] identify, on Central and Eastern European data, real economic additionality at the microeconomic level, particularly for micro-enterprises and young SMEs, with significant positive effects on the growth of beneficiary firms.

At the aggregate macroeconomic level, Honohan [20] and Beck, Klapper and Mendoza [21] show that financial depth is determined more by institutional quality than by sectoral interventions. This result converges with the estimates on governance as a long-run determinant, even though its identification remains limited by the cross-sectional dimension. More recently, the EIF report by Kraemer-Eis et al. [22] surveys analyses showing that guarantees appear effective at limiting business failures but exhibit a weaker correlation with new-lending volumes — a direct convergence with the present results.

This heterogeneity of results is explained in part by differences in methodological frameworks. Micro studies with explicit counterfactuals (Lelarge et al. [3]) can identify additionality at the firm level, but their extrapolation to the macroeconomic level is not direct if crowding-out or substitution effects exist between guaranteed and non-guaranteed loans. Macro panel studies such as the present one capture the aggregate net effect but are exposed to the inference limits documented in this article. The complementarity of the two approaches — micro for individual additionality, macro for aggregate dynamics — constitutes a research program that the present results invite others to pursue.

C. Inference with a Small Number of Clusters

The question of statistical inference with a small number of clusters has received growing attention in the econometric literature. When $N < 10$, standard clustered standard errors (HC1) are potentially underestimated because the asymptotic formulas assume $N \to \infty$. Cameron, Gelbach and Miller [5] propose the wild cluster bootstrap (WCB) as a valid inference procedure in such configurations: the dependence structure is preserved by resampling residuals at the cluster level, and a distribution of the test statistic robust to small N is reconstructed. The WCB has been shown to have good size and power properties from as few as N = 5-6 clusters, where standard asymptotic corrections fail to control size; even so, as the present design illustrates, its empirical size can depart from

the nominal level at the smallest N; in the present design it falls below nominal (empirical size ≈ 0.03; Table XXII, Appendix M).

Webb [6] refines the method by showing that Rademacher weights produce only 2^{N-1} distinct statistics for small N (that is, 64 for N = 7 — against 6^7 combinations for Webb weights), which generates coarse, staircase-shaped p-values. He recommends Webb weights — six possible values per cluster instead of two — which offer finer granularity. Wooldridge [23] (ch. 6) more generally stresses the low power of endogeneity tests such as the DWH test at small N, and Davidson and MacKinnon [24] specify the asymptotic conditions under which these tests are valid. The systematic use of the WCB with Webb weights in the SME-finance literature remains rare over the 2010-2024 period — which makes this application on aggregate European data directly comparable to the standards of Cameron et al. [5].

D. Machine Learning and Causal Identification

Mullainathan and Spiess [7] and Athey and Imbens [8] set out the fundamental distinction between predicting and identifying: minimizing out-of-sample prediction error is a different objective from estimating a conditional causal effect. Machine-learning methods are generally well suited to prediction but not to causal identification, because they optimize a loss that does not account for estimation bias. In applied economics, using them directly to infer causal effects amounts to conflating predictive correlation with causal effect.

The double machine learning of Chernozhukov et al. [9] reconciles the two by using ML to partial out the covariates nonparametrically and then estimating the causal effect through an orthogonalized moment. The central result of their framework — the orthogonality property — guarantees that the DML estimator is first-order insensitive to the estimation of the nuisance functions, which allows overparameterized ML learners to be used without biasing causal inference. Wager and Athey [25] recall that standard random forests do not correct for causal-estimation bias, and propose causal forests as an alternative — whose application on a small panel nonetheless remains little explored. The open question these works raise is whether the nonparametric flexibility of ML changes anything about the causal conclusion on such sparse data. On a small panel (N = 7, T = 15), DML can orthogonalize the covariates but cannot create statistical power where it is structurally absent — which is precisely what the results confirm.

E. Hypotheses

Three testable hypotheses follow from this review, presented below as questions directly operationalized in the empirical analysis.

H1 (Persistence): Does SME credit exhibit significantly positive autoregressive persistence, with a coefficient $\rho > 0$ and a multi-year half-life?

H2 (Governance): Does institutional quality exert a positive and significant effect on the volume of SME credit, independently of cyclical macroeconomic conditions?

H3 (Additionality): Do public guarantees exert a positive and significant additional effect on new SME lending volumes at the aggregate level?

These three hypotheses structure all the empirical results. The role of non-performing loans, which is not a tested hypothesis, is examined on an exploratory basis through Granger causality (Section IV-E): this is a secondary observation on the temporal structure of the relationships, with no claim to causal identification.

III. Data and Methodology

A. Panel and Sources

The panel covers seven European countries — Belgium, Estonia, Greece, Hungary, the Netherlands, Portugal and Slovenia — over the 2010-2024 period (15 years, 105 raw observations, 98 after creating the autoregressive lag). These countries were selected on three criteria: continuous availability of public-guarantee series in the OECD SME-finance database over the whole period, sufficient institutional heterogeneity (governance index from 0.12 for Greece to 1.84 for the Netherlands, a ratio of 1 to 15), and complete temporal coverage without interruption. This convenience panel with N = 7 is explicitly presented as exploratory: it does not claim to be representative of the European Union as a whole but illustrates SME-credit dynamics across a heterogeneous institutional architecture.

New SME lending volume (LOANNEW) and the amount of public guarantees granted (LOANGOVGUARANTEES) are drawn from the OECD Data Explorer (SME-finance Scoreboard dataset, OECD [26]). The guarantee ratio is defined as LOANGOVGUARANTEES / GDP, where GDP is nominal gross domestic product (World Bank, World Development Indicators [27]), taken for each country in the same currency as the OECD guarantee amounts, so that the ratio is a unit-free, cross-country-comparable share. The ratio measures the volume of public guarantees scaled by the size of the national economy — a synthetic indicator of the macroeconomic intensity of guarantee schemes. The interest rate spread, non-performing loans (as a percentage of total gross loans), GDP growth and the inflation rate come from the World Bank's World Development Indicators [27]. The composite governance index is drawn from the Worldwide Governance Indicators [28], synthesizing six dimensions (rule of law, regulatory quality, control of corruption, government effectiveness, political stability, and voice and accountability). The unemployment rate is included in the descriptive table for transparency but excluded from the specification: its high correlation with non-performing loans (r = +0.81) and governance (r = −0.53) would make their individual contributions statistically inextricable. Results including unemployment are available on request. Fig. 1 shows the Pearson correlation matrix among the seven panel variables.

B. Descriptive Statistics

TABLE I. DESCRIPTIVE STATISTICS (2010-2024)

Variable	N	Mean	S.D.	Min	Median	Max	CV
LOANNEW_LOG	105	22.94	1.08	21.35	23.20	25.05	0.05
Guarantee RATIO	105	0.011	0.014	0.0001	0.004	0.068	1.27
Interest spread	105	0.91	0.77	−1.26	0.83	6.00	0.84
NPL (%)	105	11.43	11.72	0.56	6.18	44.10	1.03
GDP growth (%)	105	0.37	2.67	−14.30	0.04	12.99	7.26
Inflation (%)	105	2.73	3.44	−1.74	1.97	19.40	1.26
Unemployment (%)	105	8.70	5.38	3.36	6.96	27.69	0.62
Governance	105	1.03	0.47	0.12	1.04	1.84	0.46

[a.] Variable retained for transparency, excluded from the specification (multicollinearity with NPL r = +0.81 and governance r = −0.53). The guarantee ratio is well-conditioned (CV = 1.27, skewness 1.86): the highest values belong to Portugal over 2020-2022 (ratio ≈ 0.065-0.068), and no single country dominates the distribution.

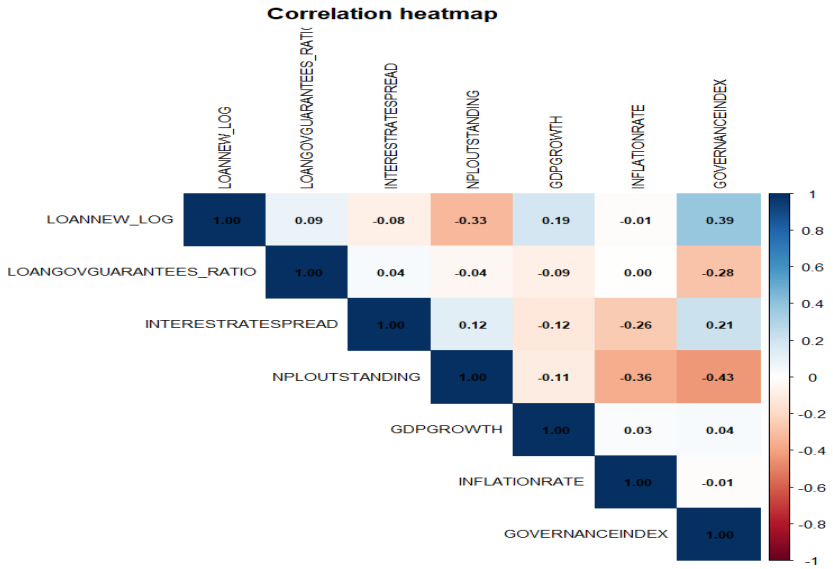

Fig. 1. Pearson correlation matrix among the 7 panel variables (7 countries, 105 observations, 2010-2024). (Note: the coefficients are shown in each cell; the blue-red gradient indicates the sign and intensity of the correlation. The correlation between LOANNEW_LOG and GOVERNANCEINDEX is the strongest for the dependent variable (r = 0.39); the strongest correlation in the whole matrix is NPL / governance (r = −0.43). The guarantee ratio (LOANGOVGUARANTEES_RATIO) exhibits moderate correlations with all variables (|r| ≤ 0.28), which validates its inclusion without any problematic multicollinearity (maximum VIF 1.62). See Section III-A and Table I (Section III-B) for the complete descriptive statistics.)

C. Main Econometric Specification

The estimated model is a dynamic panel with country and time fixed effects, on 98 observations after the lag:

$$\log(\text{LOANNEW}_{it}) = \rho \cdot \log(\text{LOANNEW}_{i,t-1}) + \beta \cdot \text{RATIO}_{it} + \gamma \cdot X_{it} + \alpha_i + \delta_t + \varepsilon_{it} \quad (1)$$

where X_{it} groups the interest rate spread, non-performing loans, GDP growth, inflation and the governance index. The country fixed effects α_i absorb all time-invariant individual heterogeneity (structural institutions, size of the banking system, credit culture). The time effects δ_t absorb common macroeconomic shocks (European financial cycle, ECB policy rate, pandemic). The autoregressive term $\log(\text{LOANNEW}_{i,t-1})$ captures the inertia of the credit market, whose presence is confirmed by all the estimators used.

The main specification uses the guarantee ratio in levels (M0). The ratio is well-conditioned (CV = 1.27, no extreme outlier), so a variance-stabilising transformation is not warranted; the ratio in levels is the most directly interpretable, its coefficient being the semi-elasticity of new lending to the guarantee-to-GDP share. Robustness to four alternative ratio specifications (99% and 95% winsorization, COVID exclusion, top-5 exclusion) is reported in Table IV and leaves the conclusion unchanged; a predetermined, lagged normalization is examined separately in Appendix L.

The dynamic fixed-effects estimator is justified by the low cross-sectional dimension (Judson and Owen [29]): at N = 7, System-GMM (Blundell and Bond [30]) produces systematic instrument proliferation (Hansen p = 1.000; Roodman [31]) and degenerate AR diagnostics (Arellano and Bond [32]; see Table VII in Appendix A, Table VIII in Appendix B for the full coefficients). The Nickell bias of the dynamic fixed-effects estimator is downward and of leading order $-(1+\rho)/(T-1)$, which at T = 14 is not negligible (about −0.12 for $\rho \approx 0.63$). A half-panel jackknife bias correction (Dhaene and Jochmans [33]) is reported in Table XXI, Appendix K; it confirms the downward direction but its magnitude cannot be pinned down

at N = 7. The FE estimate under WCB inference is retained as the main result, with ρ read as a conservative measure of persistence. The two-step GMM standard errors are corrected following Windmeijer [34]. Errors are clustered at the country level; HC1 and WCB p-values are reported side by side. The classical Hausman test opposing fixed and random effects is not estimable on this panel (7 individuals for 8 parameters: the Swamy-Arora random-effects model requires more individuals than coefficients); the choice of fixed effects therefore rests on theoretical arguments — unobserved institutional heterogeneity correlated with the regressors — corroborated by two empirical elements: cross-sectional dependence across countries (Pesaran CD test, p = 0.011) and heterogeneity of the long-run relationships, bivariate cointegration being detected in only two of seven countries (see Table XVIII, Appendix I; Tables XIX and XX, Appendix J). These elements justify absorbing the idiosyncratic trajectories through country and time fixed effects rather than pooling the observations under common coefficients.

Two methodological precautions are required regarding the stochastic properties of the series (full results in Appendix D, Fig. 6, Tables X and XI). Pesaran's cross-sectional dependence test (CD = −2.544, p = 0.011) rejects the hypothesis of independence across countries, which justifies resorting to Pesaran's CIPS test [35] rather than to the first-generation LLC (Levin, Lin and Chu [36]) and IPS (Im, Pesaran and Shin [37]) tests alone. The CIPS — the recommended standard at low N with cross-sectional dependence — settles the indeterminate cases: LOANNEW_LOG is I(1) (CIPS = −1.725), GOVERNANCEINDEX is I(1) (CIPS = −1.437), NPLOUTSTANDING is I(0) (CIPS = −2.551, significant at 5%), and LOANGOVGUARANTEES_RATIO is stationary I(0) (CIPS = −2.648, significant at 1%). GDPGROWTH also shows a slight divergence (CIPS = −1.935 vs. LLC/IPS stationary) — a minor signal with no consequence for the specification. Note: the Westerlund Group-tau and Group-alpha statistics could not be computed reliably for this panel and are reported as NA rather than interpreted. The governance index is classified as I(1) non-stationary (LLC p = 0.084; IPS p = 0.334; CIPS = −1.437; Breitung and Pesaran [38]). NPLOUTSTANDING is stationary I(0) according to CIPS — its inclusion in levels is defensible. The inclusion in levels of governance (I(1)), by contrast, raises a risk of spurious regression. The defense adopted is twofold: the fixed effects absorb the common country-specific trend, and the autoregressive term captures the dynamic adjustment — which amounts to estimating a relationship in deviation from the long-run trajectory rather than in pure levels. Consequently, the governance coefficient is interpreted as a robust directional sign, not as a precise long-run elasticity. As a robustness check against the I(1) classification, a specification with governance in first differences yields a governance coefficient that is not significant under the Webb WCB (β = +0.17, p_{WCB} = 0.33), while the persistence coefficient remains nominally significant (p_{WCB} = 0.026); the Δ-governance coefficient measures the effect of annual changes in governance — small and noisy in the WGI series — rather than of its level (Appendix N). Moreover, cointegration is ambiguous: of five available statistics, two reject the absence of cointegration (Kao p = 0.013 (one-sided test H1: α < 0); Westerlund [39] Panel-tau p = 0.030) and three do not (Engle-Granger p = 0.310; Westerlund Group-tau = NA (not reliably computable); Panel-alpha p = 0.168). This mixed result does not validate a structural long-run relationship; it is sufficient to justify the dynamic model but imposes the reservation already stated on the magnitude of the coefficients. A country-by-country bivariate ADF cointegration test (MacKinnon [40]) between LOANNEW_LOG and GOVERNANCEINDEX — the two I(1) variables — confirms this ambiguity: 2 of 7 countries show cointegration (Hungary t = −3.589 significant at 5%; Netherlands t = −3.150 significant at 10%), while the other 5 do not reject it. This cross-country heterogeneity justifies resorting to country fixed effects to absorb the idiosyncratic long-run trajectories (full results in Appendix J).

At N = 7 clusters, HC1 p-values are potentially underestimated. The wild cluster bootstrap (Cameron, Gelbach and Miller [5]) is deployed with Webb weights [6], recommended when N < 10 because Rademacher weights produce only 2^{N-1} distinct statistics. The bootstrap is stochastic (B = 9,999 random replications). The WCB multiplies the clustered residuals by random weights ($\pm\sqrt{(3/2)}$, $\pm\sqrt{(1/2)}$, ±1 for Webb weights) and recomputes the test statistic on each replication, producing a nonparametric empirical distribution of the statistic under the null hypothesis. The WCB p-value is the proportion of replications whose statistic is more extreme than the observed statistic. The procedure uses B = 9,999 replications with a fixed seed and the imposed-null variant: the residuals are recentered under H_0 before resampling, which improves empirical size at low N relative to the unrestricted variant (Cameron et al. [5]).

p_{HC1} and p_{WCB} are reported side by side in Table III; the causal interpretation rests on p_{WCB}. A coefficient with $p_{HC1} < 0.05$ but $p_{WCB} > 0.05$ is considered non-significant under valid inference. This convention is more conservative than standard practice but is the only defensible one at N = 7. Accordingly, the results on guarantees and governance are presented as indicative rather than confirmed.

D. Inference with Few Clusters: Wild Cluster Bootstrap

At N = 7 clusters, standard inference via HC1 standard errors produces rejection rates of up to 15% for a nominal size of 5% (Cameron, Gelbach and Miller [5]). The wild cluster bootstrap with Webb weights [6] is the reference inference procedure (B = 9,999 replications, imposed null). Section IV systematically presents p_{HC1} and p_{WCB} side by side; only p_{WCB} grounds the causal interpretation.

Three robustness devices complement the main estimation. The IV specification instruments the ratio with its second- and third-order lags and a pandemic interaction (COVID × $RATIO_{t-1}$), on a subsample reduced to 84 observations (2013-2024, since creating the second- and third-order instruments requires the years 2010-2012). The first stage is strong (F = 13.75 > the Stock and Yogo [41] threshold of 10), but the overidentification and residual-autocorrelation diagnostics are mixed (Sargan p = 0.063; AR(2) joint F-test p = 0.035 against an insignificant lag-2 residual coefficient, p = 0.63), so the IV estimates are treated as indicative rather than identifying. The Durbin-Wu-Hausman test points to marginal endogeneity (p =

0.076, 10% level only), with low power at N = 7 (Wooldridge [23]).

The sensitivity analysis compares five ratio specifications on the exact 98 observations of the main model. The DML PLR (Chernozhukov et al. [9]) partials out the covariates nonparametrically via a random forest (300 trees, depth 3) with temporal partitioning by country into three consecutive folds. The SHAP analysis (Lundberg and Lee [10]) compares the predictive hierarchies with and without the lagged variable in order to demonstrate mechanically the decoupling between econometric significance and predictive importance. Ridge and Lasso regressions confirm the hierarchy: in standardized terms governance is the largest coefficient (Lasso 0.42, Ridge 0.36), ahead of the guarantee ratio — Ridge/Lasso results reported in Appendix C (Fig. 4 and 5, Table IX); PanelOLS/RF comparison in Appendix G (Table XV, Fig. 8).

IV. Results

A. Dynamic Fixed-Effects Model: Main Estimator

The most stable result in Table II is persistence: $\rho = 0.629$ ($p_{WCB} = 0.065$, Section IV-B). The half-life is 1.50 years and the long-run factor $1/(1-\rho) = 2.70$ is valid only for the autoregressive term itself, not directly transposable to the other regressors. Governance is positively associated with credit ($\beta = 0.340$, HC1 $p = 0.004$). The guarantee ratio comes out positive but not significant ($\beta = +3.01$, HC1 $p = 0.339$). The other regressors — spread, NPL, GDP, inflation — do not cross the 5% threshold. This last point is unsurprising: once inertia is captured by the autoregressive term, the explanatory margin of the cyclical variables shrinks.

The within R^2 of 0.549 confirms the model's fit. RESET $p = 0.87$ and Breusch-Pagan $p = 0.21$ are favorable (valid functional form, no structural heteroskedasticity); Breusch-Godfrey $p = 0.065$ does not signal second-order residual autocorrelation at the 5% level (borderline at 10%), while Jarque-Bera $p = 0.003$ still indicates non-Gaussian residuals. As a precaution, the WCB with Webb weights — which does not assume the absence of intra-cluster autocorrelation — remains the reference inference in this article. The residual plot (Appendix D) shows a structure without systematic trend but with a few high-residual points (±0.4) at intermediate fitted values.

TABLE II. Dynamic FE Model — Specification M0 (Ratio in Levels) (Country + Time Fixed Effects, HC1-Clustered SE)

Variable	Coef. (M0)	SE_HC1	t	p_{HC1}
RATIO (β)	3.0068	3.1257	0.96	0.339
LOANNEW_LOG_LAG (ρ)	0.6293	0.1239	5.08	<0.001
GOVERNANCEINDEX	0.3401	0.1150	2.96	0.004
NPLOUTSTANDING	−0.0036	0.0025	−1.43	0.156
INTERESTRATESPREAD	−0.0189	0.0696	−0.27	0.787
GDPGROWTH	−0.0011	0.0033	−0.32	0.752
INFLATIONRATE	−0.0100	0.0172	−0.58	0.562
Within R^2	0.549	—	—	—
N obs	98	—	—	—

B. Wild Cluster Bootstrap: The Central Table

Table III is the core of the article. For each coefficient of interest, it reports the standard HC1 p-value and the Webb WCB p-value at 7 clusters. The gap between the two columns is the empirical demonstration of the pitfall of naive inference at low N. It is most visible on the two coefficients HC1 would declare significant: for governance (M0) the p-value moves from 0.018 (HC1) to 0.151 (Webb WCB), and for persistence from <0.001 to 0.065 — pushing persistence just outside the 5% threshold. The guarantee ratio, by contrast, is non-significant under both inferences (HC1 $p = 0.43$, WCB $p = 0.40$) and therefore carries no HC1→WCB gap. These gaps are not numerical artifacts: they reflect exactly what Cameron, Gelbach and Miller [5] predict theoretically for N = 7.

TABLE III. HC1 vs Webb Wild Cluster Bootstrap Inference (B = 9,999, 7 Clusters). WCB sig. is 'yes' when pWCB < 0.05. [a] pHC1 Computed on the lm-Dummies Model used for the WCB; the plm Model used elsewhere gives Slightly Different Clustered p-Values (Ratio 0.339, Governance 0.004) — the Gap Reflects Slightly Different Clustering Procedures (vcovCL vs vcovHC), with No Effect on the WCB Conclusion

Variable (spec.)	Coef.	p_{HC1}	p_{WCB} Webb	WCB sig.
RATIO raw (M0)	3.0068	0.432[a]	0.395	no
GOVERNANCE (M0)	0.340	0.018[a]	0.151	no
PERSISTENCE ρ	0.629	<0.001	0.065	no

In the main (levels) specification, no coefficient survives WCB inference at the 5% threshold; one marginal exception appears in the first-difference specification (persistence; Δ governance is not significant; Appendix N). Persistence is the only parameter approaching detectability ($\rho = 0.629$, $p_{WCB} = 0.065$, short of significance at the 5% level); guarantees and governance lie well above 5%. This does not mean their effects are null — the positive sign of the ratio is stable across the five specifications (see below), and governance keeps a positive sign throughout — but that statistical inference cannot distinguish any of them from zero with only seven clusters. This generalised non-identifiability is inherent to N = 7, not to the specification: under valid inference the panel is simply underpowered, which is itself the substantive result.

The magnitude of this power limitation can be quantified directly. A simulation on the estimated design — imposing a grid of true ratio effects, regenerating the dependent variable recursively through the autoregressive channel, and applying the Webb WCB to each simulated sample — gives the test's power as a function of the true coefficient, expressed in log points of new lending per one-standard-deviation rise in the guarantee ratio. The critical value is recalibrated by simulation to deliver a true 5% test (empirical size 0.030; calibrated threshold alpha* = 0.087). Under this size-calibrated test, the minimum effect detectable with 80% power is about 0.12 log points per standard deviation — roughly a 13% rise in new lending for a one-standard-deviation increase in the ratio. The estimated effect (about 0.047, or +5% per standard deviation, measured on the power-analysis sample) is some two to three times smaller than this threshold. This calibration is specific to the test of the guarantee ratio — a highly skewed regressor — and does not transpose to the autoregressive coefficient; the statement that no coefficient is identified at the 5% level refers to the nominal WCB in the levels specification. The failure to reject is therefore a limitation of statistical power at N = 7, not evidence that the effect is zero (full power curve in Appendix M).

Note that the pairs bootstrap reported in Table XII, Appendix E (B = 500) yields intervals that include zero for every coefficient (governance [−0.12, 0.77]; persistence [−0.04, 0.79]), consistent with the WCB conclusion of generalised non-detectability (Table XIII, Appendix F). This is expected at N = 7. Cameron et al. [5] show in simulation that the pairs bootstrap produces rejection rates of up to 15% for a nominal size of 5% at G = 5-6 clusters, and can generate pseudo-samples where the model is inestimable when the regressors vary little across clusters. The wild cluster bootstrap-t controls size far better and is the reference inference in this article; in the present design its empirical size is close to nominal (≈ 0.03 at the 5% level; Appendix M), so the nominal WCB p-values are, if anything, mildly conservative for the ratio test. The pairs bootstrap is retained in Appendix E for transparency (Fig. 7). Fig. 2 below shows the time evolution of the guarantee-ratio coefficient.

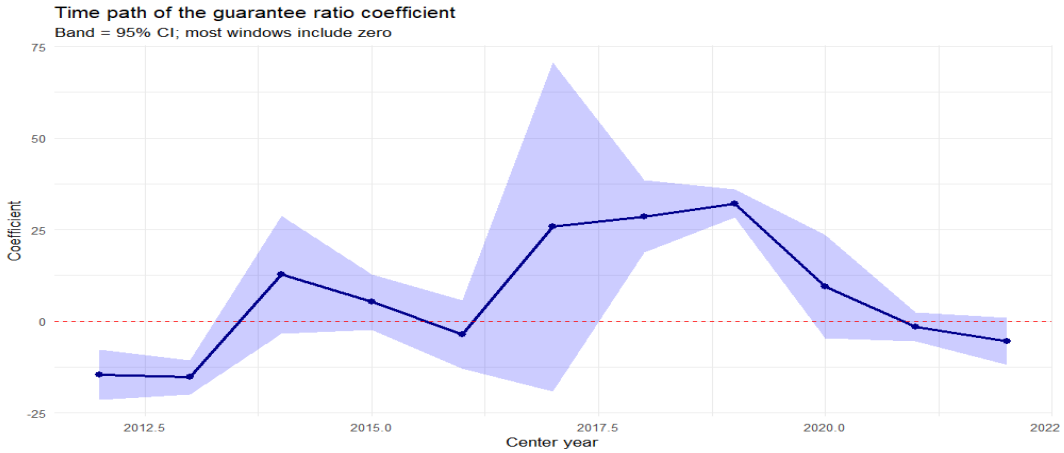

Fig. 2. Time evolution of the guarantee-ratio coefficient — rolling estimates over 5-year windows (dynamic FE, 98 observations, 7 countries, 2010-2024). (*Note: each point represents the FE coefficient estimated over a window centered on the year on the horizontal axis; the blue band is the 95% CI. The coefficient is unstable across windows and almost all intervals include zero. This temporal instability confirms the absence of a stable, identifiable effect of guarantees on aggregate credit, consistent with the WCB non-detectability (p ≈ 0.40). See Section IV-C for discussion.*)

C. Ratio Sensitivity and Robustness

Table IV reports the five ratio specifications. The Chow test on the dynamic model (mid-period break, F = 1.289, p = 0.261) does not detect structural instability — the parameters are stable over the whole 2010-2024 period, including around the pandemic. The positive sign is preserved in all five specifications, and none reaches HC1 significance (the coefficient ranges from +0.99 (excluding 2020-2021) to +3.10 (99% winsorization)), confirming at the specification level the non-detectability established under the WCB. The guarantee ratio contains no extreme outlier — its highest values (Portugal, 2020-2022) remain below 7% of GDP — so the cross-sectional result does not hinge on any single country's leverage.

The three complementary robustness checks are consistent with the non-detectability established above. System-GMM (pgmm, Blundell and Bond [30]) gives ρ = 0.771, but its diagnostics are degenerate (Hansen p = 1.000, AR(1) = AR(2) = 0.292; Roodman [31]). The pdynmc implementation returns ρ = 1.004 — an explosive value indicating model instability at N = 7. This inconsistency between the two implementations confirms that the GMM diagnostics are inconclusive on this panel (see Appendix B). System-GMM is reported for transparency only; its degenerate diagnostics preclude interpretation. The three IV estimates are sign-unstable and none is significant: the full IV (3 instruments, N = 84) gives β = −1.15 (p = 0.70), the downgraded IV β = −2.50 (p = 0.64), and the ex-COVID IV (N = 70, the seven countries all present, only 2020-2021 excluded) β = +1.51 (p = 0.78). All confidence intervals encompass zero. The IV diagnostics are mixed and are not over-interpreted. The instruments are strong in the first stage (F = 13.75 > 10), but the Sargan overidentification test rejects at the 10% level (p = 0.063), and the residual AR(2) evidence is itself divided: the lag-2 residual coefficient is insignificant (p = 0.63), whereas the joint AR(2) F-test rejects at the 5% level (F = 3.52, p = 0.035). This second signal implies possible second-order residual autocorrelation, which

would weaken RATIO_LAG2 as an instrument; combined with the order-2 Breusch-Godfrey test on the FE residuals (p = 0.065), these diagnostics caution against reading the IV estimates as more than indicative. The systematic leave-one-out analysis (Appendix J) shows that excluding Slovenia lowers ρ from 0.629 to 0.366, confirming that this country carries much of the estimated persistence. The positive sign of the ratio is preserved in all seven exclusions (β ∈ [+0.50, +4.99]) and is never significant; for the six other countries ρ stays within [0.61, 0.70]. Governance keeps a positive sign throughout but is magnitude-sensitive (β falls to 0.197, p = 0.125, when Portugal is excluded, and rises to 0.424 when Estonia is excluded).

TABLE IV. SENSITIVITY OF THE RATIO COEFFICIENT — 5 SPECIFICATIONS (★ = RETAINED AS THE MAIN ONE). NONE IS SIGNIFICANT AT 5%; THE POSITIVE SIGN IS STABLE ACROSS ALL FIVE SPECIFICATIONS

Specification	β	p_{HC1}	Sig. 5%	Within R^2
M0 raw ratio (ref.) ★	3.0068	0.339	✗	0.549
M1 winsorized 99%	3.0992	0.338	✗	0.549
M2 winsorized 95%	2.5549	0.463	✗	0.546
M3 excl. 2020-2021	0.9900	0.670	✗	0.620
M4 excl. top-5	2.5259	0.414	✗	0.565

D. Double Machine Learning and Shapley-value (SHAP) Analysis

This section puts to the test the distinction between prediction and causal identification, central to the use of machine learning in inference [7, 8]: double machine learning orthogonalizes the covariates to test whether nonparametric flexibility changes the causal conclusion, while the SHAP analysis separates predictive importance from econometric significance. The DML PLR (Table V) returns a positive point estimate on the guarantee ratio — θ_{DML} = +4.01 (SE = 1.65), 95% CI = [0.78, 7.25], nominally significant at the 5% level (p = 0.015). This significance rests on asymptotic standard errors; the DML estimate was not subjected to the cluster-robust bootstrap, and at seven clusters its asymptotic p-value warrants the same caution as the HC1 results of Section IV-A. The directly bootstrap-tested quantity is the FE ratio coefficient, which is not significant under the Webb WCB (p_{WCB} = 0.40, Section IV-B). The DML thus yields a positive point estimate but no cluster-valid evidence of an effect, consistent with the orthogonalisation logic of Chernozhukov et al. [9]: nonparametric flexibility relocates but does not create identifying power that is structurally absent at N = 7.

The SHAP analysis (Lundberg and Lee [10]) reveals a complementary methodological result that supports the discussion of the significance/predictivity decoupling. With the lag included, the guarantee ratio obtains a mean SHAP of only 0.009 (rank 5 of 7 variables); without the lag it rises to rank 1 (SHAP = 0.650). This rank reversal is a controlled illustration of a well-understood mechanism — the lagged dependent variable absorbing the predictive content otherwise loaded onto the ratio — not an additionality effect: once past inertia is in the model, the ratio carries almost no marginal predictive value. It contributes to the literature on the use of ML in panel econometrics by showing concretely how predictive salience and causal identification can diverge. Fig. 3 presents the two SHAP hierarchies side by side. Note: the SHAP values measure the marginal contribution to the random forest's prediction — they are not a causal measure. The ratio ranked first without the lag (SHAP = 0.650) is predictive in the ML model, not necessarily causal in the econometric sense: this predictive contribution partly reflects correlation with other determinants of credit, and not an isolated additionality effect.

TABLE V. DOUBLE ML PLR (CHERNOZHUKOV ET AL. [9]) — EFFECT OF THE GUARANTEE RATIO

Specification	θ_{DML}	SE	p-value	95% CI	Verdict
Raw ratio	4.013	1.651	0.015	[0.777, 7.249]	Sig. at 5% (asymptotic SE only; not cluster-bootstrapped)

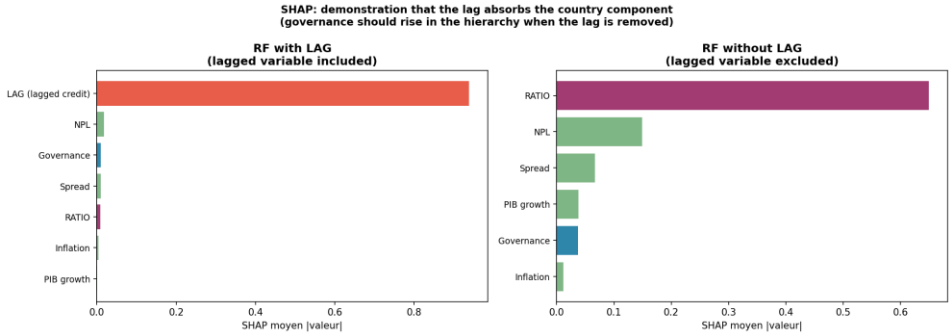

Fig. 3. Comparison of SHAP importance hierarchies with and without the lagged term LOANNEW_LOG_LAG (Random Forest, 98 observations, 7 countries, 2010-2024; Lundberg and Lee [10]). (Note: the bars represent each variable's mean absolute SHAP value — a measure of its predictive contribution to the model. With the lag (left panel), LOANNEW_LOG_LAG dominates (SHAP = 0.938) and the guarantee ratio is ranked 5th (SHAP = 0.009). Without the lag (right panel), the ratio rises to 1st rank (SHAP = 0.650). This rank jump illustrates that the autoregressive term absorbs the ratio's apparent predictive content. See Section IV-D for discussion.)

E. Granger Tests and Synthesis of the Hypotheses

The panel Granger causality test (Dumitrescu and Hurlin [42]), whose results are reported in the synthesis Table VI, clarifies the temporal structure of the relationships among the variables. Governance significantly precedes credit (Zbar = 9.15, $p < 0.001$), as do non-performing loans (Zbar = 6.28, $p < 0.001$) and inflation (Zbar = 3.22, $p = 0.001$); these Granger statistics are asymptotic and read qualitatively at N = 7. GDP growth is borderline ($p = 0.057$). The guarantee ratio, by contrast, does not precede credit (Zbar = 0.49, $p = 0.623$). A complementary panel Granger test runs the relationship in both directions and finds that new lending Granger-causes the guarantee ratio ($p = 0.008$) while the reverse does not hold ($p = 0.955$): guarantee intensity follows credit conditions rather than driving them. This pattern is consistent with contemporaneous co-determination and partial reverse causality — guarantees respond to the same macroeconomic conditions as credit, compatible with an endogenous response of guarantees to credit conditions.

The strong precedence of governance, combined with its non-detectability under the WCB in levels, may seem paradoxical. It is explained by the distinction between Granger causality (temporal precedence in the data) and statistical identification under robust inference (significance of the conditional parameter in the fixed-effects model). Governance precedes credit because it captures long-run institutional trends that inform future credit levels. Its FE coefficient is positive and its sign robust. But at N = 7, its magnitude cannot be distinguished from zero under valid inference in levels — which does not call Granger causality into question, but underscores that precise quantification of the effect requires more clusters.

TABLE VI. STATUS OF THE THREE HYPOTHESES AND THE SECONDARY OBSERVATION (WEBB WCB INFERENCE)

	Key test/result	Status
H1 — Credit persistence	$\rho = 0.629$; $p_{WCB} = 0.065$	Closest to detectable; not significant at 5% in levels ($p_{WCB} = 0.065$); nominally significant in first differences ($p_{WCB} = 0.026$)
H2 — Governance as LR determinant	$\beta = 0.340$; $p_{WCB} = 0.151$; Granger $p < 0.001$	Indicative — not detectable in levels or in first differences ($p_{WCB} = 0.33$)
H3 — Guarantees (additionality)	Positive sign 5/5; FE $p_{WCB} = 0.40$; DML $p = 0.015$ (asymptotic SE only, not bootstrapped)	Indicative — not detectable under WCB
Obs. — NPL indirect channel (exploratory)	Granger $p < 0.001$	Suggestive — Granger precedence, no causal identification
Obs. — Reverse causality (new lending → guarantees)	Granger $p = 0.008$; reverse $p = 0.955$ (asymptotic; read qualitatively at N = 7)	New lending precedes guarantee intensity

V. DISCUSSION

A. Persistence: The Least Fragile Result

Persistence $\rho \approx 0.63$ is the least fragile result of this article. The FE benchmark (0.629) is corroborated by the IV 2SLS estimate (0.622, on 84 obs); the System-GMM value (0.771) rests on degenerate diagnostics and is not used to corroborate it (Appendix B). Under valid WCB inference it is the only coefficient that approaches detectability ($p_{WCB} = 0.065$, short of significance at the 5% level; the WCB's finite-sample behaviour at seven clusters is examined in Appendix M); in a first-difference specification it is nominally significant ($p_{WCB} = 0.026$; Appendix N). A reservation is in order: the leave-one-out analysis shows that excluding Slovenia lowers ρ to 0.366, so the estimated persistence is partly carried by this country; for the six remaining countries $\rho \in [0.61, 0.70]$ — still substantial but more moderate. The half-panel jackknife bias correction (Appendix K) points to downward Nickell bias but is itself imprecise at N = 7. Its half-life of 1.50 years means that a shock to SME credit dissipates over several years. This inertia has a direct normative implication: public policies targeting SME credit must reason over a multi-year horizon and not evaluate their effectiveness in the short term. A scheme activated in response to a shock produces most of its effects over the following two to three years, not within the current budget year.

Persistence is the only parameter that approaches detectability, which distinguishes it from guarantees and governance, whose signs are consistent but which remain undetectable under WCB in the main specification. This distinction — one near-identifiable parameter, two indicative ones — delimits what the panel can establish. With N = 7 countries and T = 15 years, the design does not have the statistical power needed to detect a moderate effect of guarantees on aggregate credit — which is itself informative for anyone undertaking a similar evaluation with finer data.

B. Guarantees and Governance: Indicative, Not Identified

On guarantees, no estimator supports a firm conclusion. The sign is positive across all five FE specifications but never significant under WCB inference ($p = 0.40$); the DML returns a positive point estimate ($\theta = +4.0$) significant only under asymptotic standard errors, which, at seven clusters, are not reliable; it was not subjected to the cluster-robust bootstrap. The three IV estimates are sign-unstable (−1.15, −2.50, +1.51) and none is significant. The ratio does not Granger-cause credit. Across methods and signs, no specification identifies an effect: what is established is not the absence of an effect but the impossibility of deciding at N = 7.

This picture does not, however, invalidate the additionality hypothesis. It says rather that, at this level of aggregation and with this panel architecture, additionality is not separable from the credit's own dynamics. The most plausible economic reading is one of partial reverse causality: the panel Granger

test indicates that new lending precedes guarantee intensity ($p = 0.008$) rather than the reverse ($p = 0.955$), consistent with guarantee intensity responding to credit conditions. The ratio × GDP interaction ($p = 0.89$) does not confirm an explicit countercyclical mechanism, so the design cannot separate reverse causality from an aggregate substitution between guaranteed and non-guaranteed credit — which is precisely why a micro level of analysis would be needed.

Governance exhibits a profile similar to that of guarantees: a positive sign throughout (FE $\beta = 0.340$), strong temporal precedence in the Granger sense ($p < 0.001$, asymptotic), but $p_{WCB} = 0.151$. Granger precedence and WCB non-detectability are not contradictory: Granger captures the temporal relationship in the raw data, whereas the WCB tests whether the model's conditional parameter can be distinguished from zero at seven clusters. The power of the WCB at $N = 7$ for an effect of this magnitude is likely insufficient — which means that the failure to reject H0 is as compatible with a true effect the design lacks the power to detect as with a genuine absence of effect.

The co-orientation of the ratio and non-performing loans on the first PCA axis supports this stress-driven reading. The DWH test points to marginal endogeneity ($p = 0.076$, 10% only), with limited power at $N = 7$. This is consistent with the conclusions of Kraemer-Eis et al. [22]: at the aggregate level, guarantee intensity appears to respond to credit conditions rather than to stimulate new lending. A size-calibrated power analysis (Appendix M) further shows the design could not reliably detect effects smaller than about 13% per standard deviation — more than twice the estimated effect — so the failure to identify an effect of guarantees reflects limited power at $N = 7$, not an established null. The conclusion is identical for governance: the positive sign is retained as robust ($\beta = 0.340$) and defensible, but not the magnitude — presenting a precise long-run elasticity for a parameter indistinguishable from zero in levels at $N = 7$ would be inconsistent with the non-identifiability thesis.

C. Two Methodological Contributions: Robust Inference and the Prediction-identification Decoupling

The gap between p_{HC1} and p_{WCB} is marked on the coefficients that standard inference would declare significant: governance moves from 0.018 to 0.151, and persistence from <0.001 to 0.065. This gap matches what Cameron, Gelbach and Miller [5] and Webb [6] establish theoretically for a small number of clusters. Documenting it on an SME-finance panel provides an empirical illustration of this phenomenon and underscores that results resting on HC1 standard errors in small panels are better interpreted with caution.

The second methodological contribution belongs to machine learning applied to causal identification, and the SHAP demonstration without the lag is its central illustration. The guarantee ratio moves from rank 5 to rank 1 when the autoregressive lag is removed — the exact mechanism predicted by the theory of dynamic models: the lagged dependent variable absorbs the predictive content otherwise loaded onto the ratio. This is a controlled illustration of a well-understood mechanism, not a discovery, and contributes to the literature on the use of ML in panel econometrics by showing concretely how predictive salience and causal identification can diverge.

D. Normative Implications

The gap between a statistically detectable effect and genuine additionality is the most direct normative consequence of these results. A scheme whose aggregate effect on new lending is undetectable under robust inference is not thereby shown to be ineffective: the data are equally consistent with a modest effect the panel lacks the power to detect and with guarantees responding endogenously to credit conditions, and distinguishing the two would require finer data. Measuring the success of guarantees by the volume of credit covered alone would be to use the wrong evaluation criterion. The OECD [1] guidance (pp. 17-18) on improving the reporting of guaranteed loans, and that of the World Bank (Carvajal and Didier [2]) on the design principles of guarantee funds, fall within this logic. These results concern the aggregate effect at the country level. They say nothing about the effect an individual guarantee may have on a given borrower — a question that only microeconomic data could settle. The design guidance operates precisely at this micro level; the aggregate analysis here complements it without contradicting it.

The design sequencing identified by Carvajal and Didier [2] — legal framework, sustainable financing, prudential supervision, transparent governance — constitutes a useful normative framework, distinct from a recommendation to expand guaranteed volumes. These results do not support a recommendation to expand or reduce guarantees — they document that their aggregate effect is not separable from the autoregressive dynamics of credit at $N = 7$. What these results do support is more targeted: governance precedes credit (Granger $p < 0.001$, asymptotic). For policymakers, this argues for microeconomic evaluations with counterfactuals — rather than aggregate volume indicators — as the primary criterion for the effectiveness of guarantee schemes.

VI. Conclusion

This article documents three results on SME-credit dynamics in a European panel of seven countries. The least fragile result is persistence: $\rho = 0.629$, a half-life of 1.50 years, the only parameter approaching detectability under the wild cluster bootstrap ($p_{WCB} = 0.065$, short of significance at the 5% level). Public guarantees and institutional governance exhibit signs consistent with the hypotheses but remain statistically undetectable in the main specification under valid inference with seven clusters; the double machine learning returns a positive coefficient on the ratio that is significant only under asymptotic standard errors, which are not reliable at seven clusters and were not subjected to the cluster-robust bootstrap. No coefficient is robustly identified at the 5% level — a generalised non-identifiability that is not a failure but precisely what a low-N exploratory panel can honestly establish; a size-calibrated power analysis confirms that the design could not reliably detect effects smaller than about 13% per standard deviation — more than twice the estimated effect — so this reflects the panel's limited power rather than evidence of no effect.

The central methodological contribution is the empirical documentation of the HC1/WCB gap on real data. The standard p-values rise sharply as soon as inference is corrected for N = 7 clusters — a result expected theoretically but rarely illustrated on an SME-finance panel. Reporting the two side by side (Table III) makes explicit the over-interpretation to which small panels are exposed.

Several limitations bound the conclusions. The cross-sectional dimension (N = 7) constrains statistical power for subtle effects. The absence of firm-level data prevents an individual measurement of additionality. At seven clusters, individual countries exert large leverage on the estimates — excluding Slovenia lowers the persistence coefficient from 0.63 to 0.37 — a fragility inherent to the panel architecture.

The logical continuation of this work first requires enlarging the panel. Fifteen to twenty European countries over the same period would provide the statistical power needed to distinguish the effects of governance and guarantees under the WCB. That is the most direct extension. A second angle would exploit the national regulatory discontinuities documented by the OECD and the ECB — changes in caps, coverage rates, eligibility conditions: these are exogenous causal instruments that the present design could not mobilize. Finally, firm-bank matching would allow additionality to be measured at the individual level. The aggregate level, structurally, cannot do so.

DECLARATIONS

FUNDING

This research received no specific funding from the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST STATEMENT

The authors declare no financial or personal conflict of interest that could have influenced this work.

DATA AND CODE AVAILABILITY STATEMENT

The panel data (seven European countries, 2010-2024) come from public sources (OECD, World Bank).

GENERATIVE AI USE STATEMENT

The authors used a generative AI assistant for formatting, language editing, and help with phrasing.

REFERENCES

[1] OECD, Financing SMEs and Entrepreneurs 2024: An OECD Scoreboard. Paris, France: OECD Publishing, 2024. [Online]. Available: https://doi.org/10.1787/fa521246-en

[2] A. Carvajal and T. Didier, Boosting SME Finance for Growth: The Case for More Effective Support Policies. Washington, DC, USA: World Bank, 2024. [Online]. Available: https://documents.worldbank.org

[3] C. Lelarge, D. Sraer, and D. Thesmar, "Entrepreneurship and credit constraints: Evidence from a French loan guarantee program," in International Differences in Entrepreneurship, J. Lerner and A. Schoar, Eds. Chicago, IL, USA: Univ. Chicago Press, 2010, pp. 243–273.

[4] M. Boschi, A. Girardi, and M. Ventura, "Partial credit guarantees and SMEs financing," J. Financial Stability, vol. 15, pp. 182–194, 2014.

[5] A. C. Cameron, J. B. Gelbach, and D. L. Miller, "Bootstrap-based improvements for inference with clustered errors," Rev. Econ. Statist., vol. 90, no. 3, pp. 414–427, 2008.

[6] M. D. Webb, "Reworking wild bootstrap based inference for clustered errors," Queen's Univ., Kingston, ON, Canada, Working Paper, 2013.

[7] S. Mullainathan and J. Spiess, "Machine learning: An applied econometric approach," J. Econ. Perspect., vol. 31, no. 2, pp. 87–106, 2017.

[8] S. Athey and G. Imbens, "Machine learning methods that economists should know about," Annu. Rev. Econ., vol. 11, pp. 685–725, 2019.

[9] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," Econometrics J., vol. 21, no. 1, pp. C1–C68, 2018.

[10] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 4765–4774.

[11] J. E. Stiglitz and A. Weiss, "Credit rationing in markets with imperfect information," Amer. Econ. Rev., vol. 71, no. 3, pp. 393–410, 1981.

[12] H. Bester, "Screening vs. rationing in credit markets with imperfect information," Amer. Econ. Rev., vol. 75, no. 4, pp. 850–855, 1985.

[13] D. Besanko and A. Thakor, "Collateral and rationing: Sorting equilibria in monopolistic and competitive credit markets," Int. Econ. Rev., vol. 28, no. 3, pp. 671–689, 1987.

[14] A. N. Berger and G. F. Udell, "A more complete conceptual framework for SME finance," J. Banking Finance, vol. 30, no. 11, pp. 2945–2966, 2006.

[15] A. W. A. Boot, "Relationship banking: What do we know?," J. Financial Intermediation, vol. 9, no. 1, pp. 7–25, 2000.

[16] M. A. Petersen and R. G. Rajan, "The benefits of lending relationships: Evidence from small business data," J. Finance, vol. 49, no. 1, pp. 3–37, 1994.

[17] L. Cusmano, "SME and entrepreneurship financing: The role of credit guarantee schemes and mutual guarantee societies in supporting finance for small and medium-sized enterprises," OECD SME Entrepreneurship Papers, no. 1, 2018.

[18] J. C. Gozzi and S. Schmukler, "Public credit guarantees and access to finance," Univ. Warwick, Dept. Econ., Warwick Econ. Res. Papers, no. 1122, 2016.

[19] P. Asdrubali and S. Signore, "The economic impact of EU guarantees on credit to SMEs," European Commission, European Economy Discussion Papers, no. 002, 2015.

[20] P. Honohan, "Partial credit guarantees: Principles and practice," J. Financial Stability, vol. 6, no. 1, pp. 1–9, 2010.

[21] T. Beck, L. Klapper, and J. Mendoza, "The typology of partial credit guarantee funds around the world," J. Financial Stability, vol. 6, no. 1, pp. 10–25, 2010.

[22] H. Kraemer-Eis, A. Botsari, S. Gvetadze, F. Lang, and W. Torfs, "The European small business finance outlook 2023," European Investment Fund, EIF Working Paper 2023/96, 2023. [Online]. Available: https://www.eif.org/news_centre/publications/eif_working_paper_2023_96.pdf

[23] J. M. Wooldridge, Econometric Analysis of Cross Section and Panel Data, 2nd ed. Cambridge, MA, USA: MIT Press, 2010.

[24] R. Davidson and J. G. MacKinnon, "Testing for consistency using artificial regressions," Econometric Theory, vol. 5, no. 3, pp. 363–384, 1989.

[25] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," J. Amer. Statist. Assoc., vol. 113, no. 523, pp. 1228–1242, 2018.

[26] OECD, "Financing SMEs and Entrepreneurs — Scoreboard," OECD Data Explorer, dataset DSD_SMEE_FINANCING, OECD Publishing, Paris, 2026. [Online]. Available: https://data-explorer.oecd.org Accessed: Jan. 2026.

[27] World Bank, "World Development Indicators," World Bank, Washington, DC, USA, 2026. [Online]. Available: https://databank.worldbank.org/source/world-development-indicators Accessed: Jan. 2026.

[28] D. Kaufmann, A. Kraay, and M. Mastruzzi, "The worldwide governance indicators: Methodology and analytical issues," Hague J. Rule Law, vol.

3, no. 2, pp. 220–246, 2011. [Online]. Available: https://www.worldbank.org/en/publication/worldwide-governance-indicators Accessed: Jan. 2026.

[29] R. A. Judson and A. L. Owen, "Estimating dynamic panel data models: A guide for macroeconomists," Econ. Lett., vol. 65, no. 1, pp. 9–15, 1999.

[30] R. Blundell and S. Bond, "Initial conditions and moment restrictions in dynamic panel data models," J. Econometrics, vol. 87, no. 1, pp. 115–143, 1998.

[31] D. Roodman, "A note on the theme of too many instruments," Oxford Bull. Econ. Statist., vol. 71, no. 1, pp. 135–158, 2009.

[32] M. Arellano and S. Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," Rev. Econ. Stud., vol. 58, no. 2, pp. 277–297, 1991.

[33] G. Dhaene and K. Jochmans, "Split-panel jackknife estimation of fixed-effect models," Rev. Econ. Stud., vol. 82, no. 3, pp. 991–1030, 2015.

[34] F. Windmeijer, "A finite sample correction for the variance of linear efficient two-step GMM estimators," J. Econometrics, vol. 126, no. 1, pp. 25–51, 2005.

[35] M. H. Pesaran, "A simple panel unit root test in the presence of cross-section dependence," J. Appl. Econometrics, vol. 22, no. 2, pp. 265–312, 2007.

[36] A. Levin, C. F. Lin, and C. S. J. Chu, "Unit root tests in panel data: Asymptotic and finite-sample properties," J. Econometrics, vol. 108, no. 1, pp. 1–24, 2002.

[37] K. S. Im, M. H. Pesaran, and Y. Shin, "Testing for unit roots in heterogeneous panels," J. Econometrics, vol. 115, no. 1, pp. 53–74, 2003.

[38] J. Breitung and M. H. Pesaran, "Unit roots and cointegration in panels," in The Econometrics of Panel Data, L. Matyas and P. Sevestre, Eds. Berlin, Germany: Springer, 2008, pp. 279–322.

[39] J. Westerlund, "Testing for error correction in panel data," Oxford Bull. Econ. Statist., vol. 69, no. 6, pp. 709–748, 2007.

[40] J. G. MacKinnon, "Critical values for cointegration tests," in Long-Run Economic Relationships: Readings in Cointegration, R. F. Engle and C. W. J. Granger, Eds. Oxford, U.K.: Oxford Univ. Press, 1991, pp. 267–276.

[41] J. H. Stock and M. Yogo, "Testing for weak instruments in linear IV regression," in Identification and Inference for Econometric Models, D. W. K. Andrews and J. H. Stock, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2005, pp. 80–108.

[42] E.-I. Dumitrescu and C. Hurlin, "Testing for Granger non-causality in heterogeneous panels," Econ. Model., vol. 29, no. 4, pp. 1450–1460, 2012.

Methodological Appendices

A. *Difference-GMM: Full Coefficients (Invalid, Reported for Transparency)*

TABLE VII. Difference-GMM Coefficients (Arellano and Bond [32]). Estimation on 91 Observations (7 Countries, 2012-2024 after Double Differencing). The Difference-GMM is Invalid at N = 7: Across Instrument Depths p is unstable and Mostly Aberrant (−1.35 at lag 2:2, −2.36 at Lag 2:3, 0.18 Only at Lag 2:4), a Value Inside (0,1) Appearing Only at the Longest Depth, so no Depth is Defensible as a Preferred Specification. Reported for Transparency Only; the Coefficients Shown are for Lag 2:4

Variable	Coef.	Std. err.	t	p	Status
LOANNEW_LOG (t−1)	0.180	1.709	0.105	0.916	Invalid
LOANGOVGUARANTEES_RATIO	18.134	38.464	0.471	0.637	Invalid
GDPGROWTH	−0.0348	0.0559	−0.623	0.533	Invalid
INFLATIONRATE	−0.0101	0.0074	−1.367	0.172	Invalid
NPLOUTSTANDING	0.0757	0.0854	0.887	0.375	Invalid
GOVERNANCEINDEX	−0.0876	2.437	−0.036	0.971	Invalid

B. *System-GMM: Full Diagnostics*

TABLE VIII. System-GMM Coefficients (Blundell and Bond [30]; Windmeijer [34]). p pgmm = 0.771 (on Degenerate Diagnostics; Not Interpreted); p pdynmc = 1.004 (Explosive). Instrument Proliferation; Hansen p = 1.000 (Roodman [31]); AR(1) = AR(2) = 0.292 (Degenerate)

Variable	Coef.	Std. err.	t	p	Status
LOANNEW_LOG (t−1)	0.771	0.139	5.536	<0.001	Indicative
LOANGOVGUARANTEES_RATIO	1.054	23.273	0.045	0.964	Indicative
GDPGROWTH	−0.0568	0.0432	−1.314	0.189	Indicative
INFLATIONRATE	0.0061	0.0229	0.268	0.789	Indicative
NPLOUTSTANDING	−0.0056	0.0424	−0.133	0.894	Indicative
GOVERNANCEINDEX	6.388	4.774	1.338	0.181	Indicative

C. Principal Component Analysis and Penalized Regressions

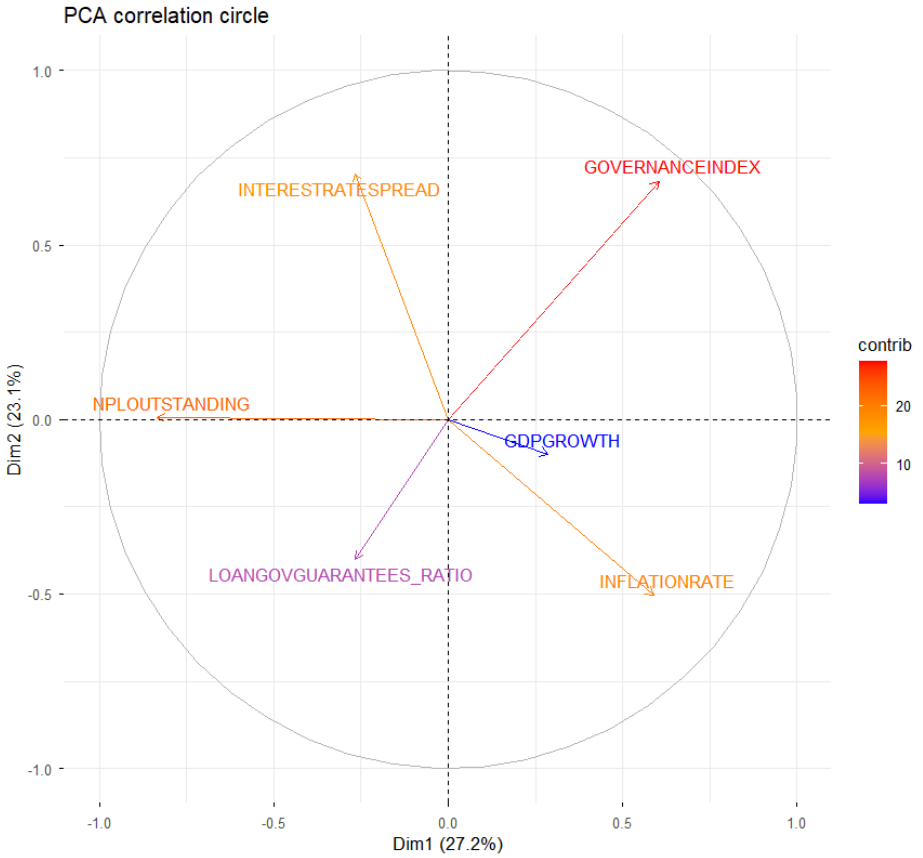

Fig. 4. PCA correlation circle — 6 regressors of the main model (excluding the dependent variable and the lag), 7 countries, 2010-2024.

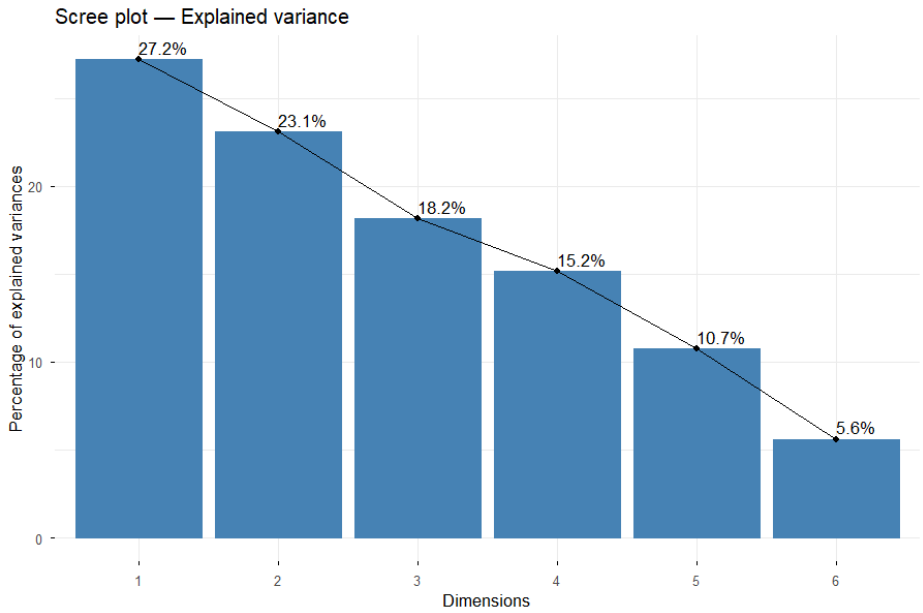

Fig. 5. Scree plot — variance explained by principal component. (Note: the first two components explain 50.3% of the variance (27.2% and 23.1%). The guarantee ratio and non-performing loans load on the negative side of the first axis (financial stress), opposite to the governance index, while interest spread and governance load positively on the second axis.)

TABLE IX. STANDARDIZED RIDGE AND LASSO COEFFICIENTS (COEFFICIENT × STANDARD DEVIATION, COMPARABLE ACROSS REGRESSORS). GOVERNANCE IS THE LARGEST STANDARDIZED COEFFICIENT (LASSO 0.42, RIDGE 0.36), AHEAD OF THE GUARANTEE RATIO — A PREDICTIVE HIERARCHY CONSISTENT WITH THE FE

Variable	Ridge	Lasso
(Constant)	—	—
LOANGOVGUARANTEES_RATIO	0.174	0.203
INTERESTRATESPREAD	−0.135	−0.152
NPLOUTSTANDING	−0.164	−0.134
GDPGROWTH	0.149	0.154
INFLATIONRATE	−0.092	−0.079
GOVERNANCEINDEX	0.356	0.420

D. Panel Unit-Root Tests and Detailed Fixed-Effects Diagnostics

TABLE X. PANEL UNIT-ROOT TESTS LLC AND IPS, COMPLEMENTED BY CIPS (PESARAN) — 7 VARIABLES, 7 COUNTRIES, 2010-2024

Variable	LLC	p LLC	IPS	p IPS	CIPS	Final conclusion
LOANNEW_LOG	−1.802	0.036	−1.069	0.143	−1.725	Non-stationary I(1)
LOANGOVGUARANTEES_RATIO	−2.278	0.011	−2.044	0.021	−2.648	Stationary I(0)
INTERESTRATESPREAD	−3.018	0.001	−2.653	0.004	−2.604	Stationary I(0)
NPLOUTSTANDING	−1.850	0.032	−0.575	0.283	−2.551	Stationary I(0)
GDPGROWTH	−3.466	<0.001	−3.247	<0.001	−1.935	Stationary I(0)
INFLATIONRATE	−4.876	<0.001	−3.557	<0.001	−5.383	Stationary I(0)
GOVERNANCEINDEX	−1.378	0.084	−0.429	0.334	−1.437	Non-stationary I(1)

TABLE XI. MAIN FE MODEL DIAGNOSTICS

Test	Statistic	p	Conclusion
RESET (dynamic)	0.027	0.871	Valid functional form
Breusch-Pagan	9.699	0.206	No structural heteroskedasticity
Jarque-Bera	11.921	0.003	Non-normal residuals
Breusch-Godfrey	5.483	0.065	No second-order autocorrelation at 5% (borderline at 10%)

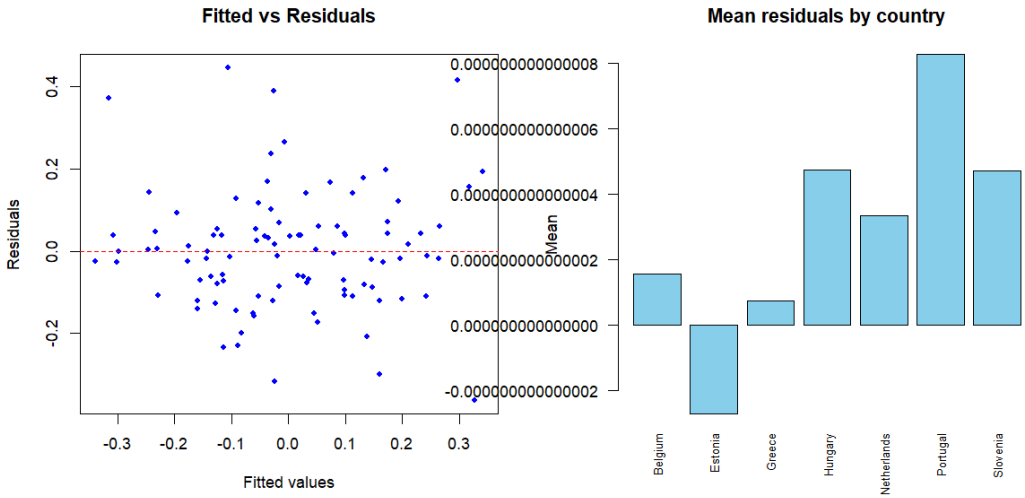

Fig. 6. Residual plot (fitted values vs residuals; country means).

E. Country-Cluster Bootstrap: Detailed Statistics (B = 500)

TABLE XII. COUNTRY-CLUSTER BOOTSTRAP CONFIDENCE INTERVALS (B = 500 REPLICATIONS, RESAMPLING OF WHOLE CLUSTERS). 95% CI: GOVERNANCE [−0.115, 0.772] INCLUDES ZERO; PERSISTENCE [−0.035, 0.789] INCLUDES ZERO

Variable	Mean	Median	Std. dev.	2.5% CI	97.5% CI
LOANGOVGUARANTEES_RATIO	3.025	2.668	5.392	−6.362	16.255
INTERESTRATESPREAD	−0.027	−0.019	0.091	−0.212	0.084
NPLOUTSTANDING	−0.005	−0.003	0.007	−0.019	0.006
GDPGROWTH	0.347	0.000	0.971	−1.164	3.353
INFLATIONRATE	−0.010	−0.011	0.025	−0.071	0.034
GOVERNANCEINDEX	0.365	0.376	0.208	−0.115	0.772
LOANNEW_LOG_LAG	0.533	0.610	0.218	−0.035	0.789

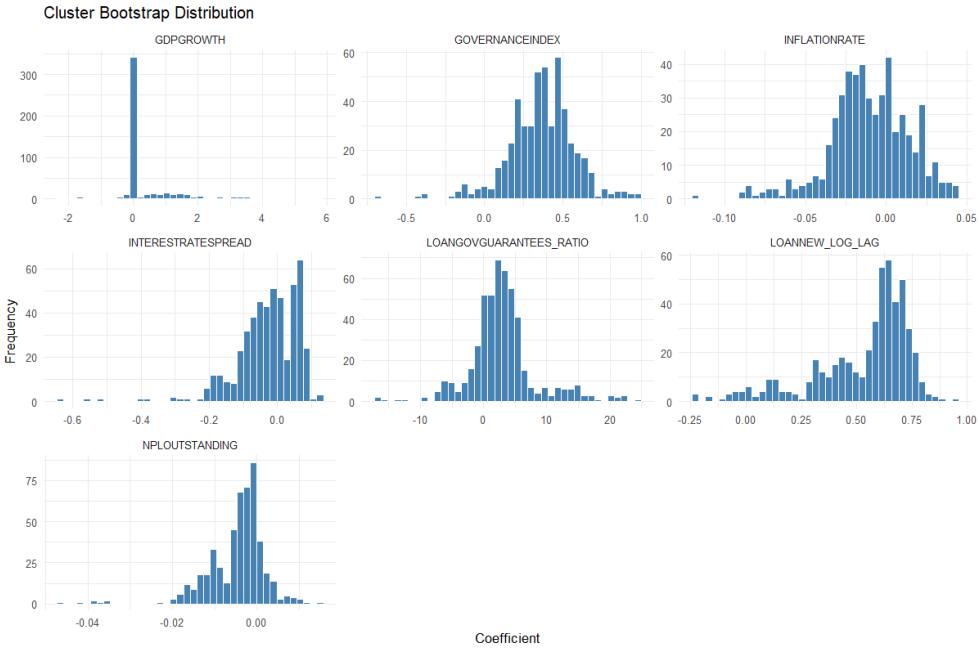

Fig. 7. Bootstrap distributions by variable.(Note: the pairs bootstrap is unreliable at N = 7 — it produces excessively high rejection rates with a small number of clusters — which is why the Webb WCB is the reference inference; these intervals are reported for transparency only.)

F. Test of H3 — Ratio × GDP-Growth Interaction (Countercyclical Effect): Details

TABLE XIII. TEST OF H3 — RATIO × GDP-GROWTH INTERACTION. INTERACTION COEFFICIENT P = 0.894 (NOT SIGNIFICANT): THE COUNTERCYCLICAL DIMENSION OF H3 (ADDITIONALITY) IS NOT CONFIRMED

Variable	Coef.	Std. err.	t	p
LOANGOVGUARANTEES_RATIO	3.007	3.146	0.956	0.342
INTERESTRATESPREAD	−0.0184	0.0707	−0.260	0.796
NPLOUTSTANDING	−0.0035	0.0025	−1.411	0.163
GDPGROWTH	−0.0022	0.0086	−0.258	0.797
INFLATIONRATE	−0.0100	0.0172	−0.583	0.562
GOVERNANCEINDEX	0.342	0.117	2.922	0.005
LOANNEW_LOG_LAG	0.629	0.124	5.063	<0.001
RATIO × GDPGROWTH	1.063	7.941	0.134	0.894

G. *Random Forest — Temporal Validation and Panel OLS Versus Random Forest Comparison*

TABLE XIV. PANELOLS VS RANDOM FOREST COMPARISON: PANELOLS COEFFICIENT AND SIGNIFICANCE, SIGN, AND RANDOM-FOREST PERMUTATION IMPORTANCE

Variable	PanelOLS coef.	p	Sig. 5%	Sign	RF import. (perm.)
LOANNEW_LOG_LAG	0.629	<0.001	Yes	+	1.220
LOANGOVGUARANTEES_RATIO	3.007	0.398	No	+	0.082
GDPGROWTH	−0.0011	0.780	No	−	0.011
INTERESTRATESPREAD	−0.0189	0.811	No	−	0.015
INFLATIONRATE	−0.0100	0.608	No	−	0.0005
NPLOUTSTANDING	−0.0036	0.210	No	−	−0.00004
GOVERNANCEINDEX	0.340	0.011	Yes	+	−0.008

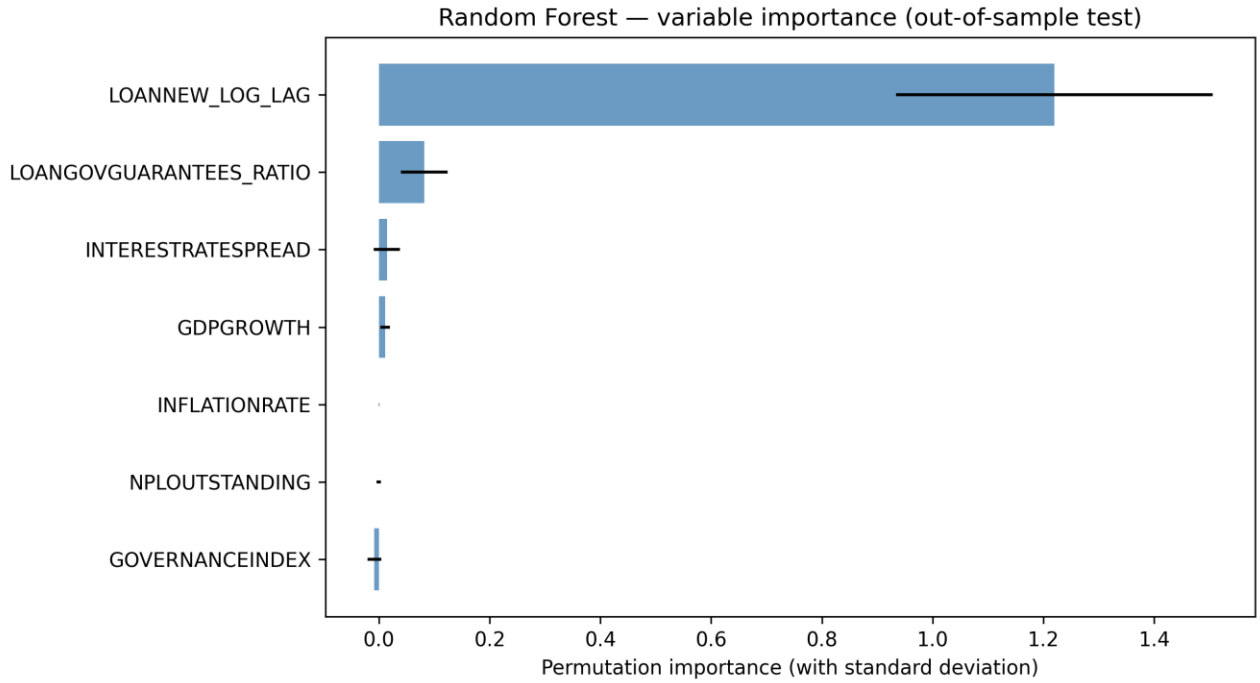

Fig. 8. Variable importance from the random forest (Gini impurity); the lag and governance rank highest. Gini-impurity importance differs from the permutation importance of Table 14, which ranks governance lower.(Note: the random forest optimizes prediction, not causal identification (Mullainathan and Spiess [7]; Wager and Athey [25]).)

H. *Instrumental Variables and Durbin-Wu-Hausman: Technical Details, Residual Autocorrelation, Downgraded and Non-Pandemic Specifications*

TABLE XV. FULL IV 2SLS RESULTS (REFERENCE SPECIFICATION, 84 OBSERVATIONS)

Variable	IV coef.	Std. err.	t	p
LOANGOVGUARANTEES_RATIO	−1.145	2.975	−0.385	0.702
LOANNEW_LOG_LAG	0.622	0.133	4.672	<0.001
INTERESTRATESPREAD	−0.0208	0.0566	−0.367	0.715
NPLOUTSTANDING	−0.0062	0.0032	−1.963	0.054
GDPGROWTH	0.0013	0.0041	0.331	0.742
INFLATIONRATE	−0.0006	0.0139	−0.042	0.966
GOVERNANCEINDEX	0.297	0.125	2.377	0.021

TABLE XVI. IV DIAGNOSTICS: INSTRUMENT STRENGTH, OVERIDENTIFICATION, AND RESIDUAL AUTOCORRELATION

Diagnostic	Value
First-stage F (Stock-Yogo [41])	13.75 ($p < 0.001$)
Instrument strength	Strong ($F > 10$)
Sargan test (overidentification)	5.536 ($p = 0.063$)
Sargan verdict	Borderline (rejects at 10%)
Residual AR(2) (F)	3.518 ($p = 0.035$)
Resid. coef. (t−2)	0.059 ($p = 0.628$)
AR(2) verdict	Mixed: lag-2 coef. n.s. ($p = 0.63$); joint F rejects at 5% ($p = 0.035$)

TABLE XVII. DURBIN-WU-HAUSMAN TEST: MARGINAL ENDOGENEITY OF THE RATIO (P = 0.076). DAVIDSON AND MACKINNON [24] RECALL THE TEST'S LOW POWER AT N = 7. ROBUSTNESS SPECIFICATIONS: DOWNGRADED IV (2 INSTR.) B = −2.50 (P = 0.640); EX-2020-2021 IV (N = 70) B = +1.51 (P = 0.784)

Diagnostic	Value
Wu-Hausman test (endogeneity)	$F = 3.272$ ($p = 0.076$)
Verdict	Marginal endogeneity (10% only)
FE coefficient	3.007 (SE 3.126)
Full IV coefficient	−1.145 (SE 5.186, Hausman auxiliary; cf. 2.975 in Table 15)
FE → IV change	−4.152 (sign reversal)

I. *Double Machine Learning (Partially Linear Model): Temporal Folds, Random-Forest Hyperparameters, Full Results*

TABLE XVIII. FULL DML PLR RESULTS — A SINGLE SPECIFICATION (M0). RF HYPERPARAMETERS: 300 TREES, DEPTH 3, 3 TEMPORAL FOLDS PER COUNTRY (CHERNOZHUKOV ET AL. [9]).

Specification	θ_{DML}	Std. err.	p	95% CI	Sig. 5%
Raw ratio (M0)	4.013	1.651	0.015	[0.777, 7.249]	Yes (asymptotic SE only; not bootstrapped)

J. *Leave-One-Out and Country-by-Country Bivariate Cointegration*

TABLE XIX. LEAVE-ONE-OUT — INFLUENCE OF EACH COUNTRY (DYNAMIC FE, SPECIFICATION M0). POSITIVE SIGN OF BRATIO PRESERVED 7/7. SLOVENIA CARRIES MUCH OF THE MAGNITUDE OF P

Excluded country	N	ρ	β ratio	p ratio	β gov.	p gov.
None (ref. M0)	98	0.629	3.007	0.339	0.340	0.004
Belgium	84	0.630	2.873	0.338	0.385	0.002
Estonia	84	0.655	3.300	0.272	0.424	0.004
Greece	84	0.653	0.503	0.865	0.312	0.008
Hungary	84	0.665	2.490	0.373	0.359	0.002
Netherlands	84	0.703	1.448	0.666	0.349	0.003
Portugal	84	0.611	4.051	0.475	0.197	0.125
Slovenia	84	0.366	4.992	0.242	0.499	<0.001

TABLE XX. BIVARIATE COINTEGRATION LOANNEW_LOG / GOVERNANCEINDEX BY COUNTRY (ADF ON OLS RESIDUALS, MACKINNON [40], 5% CV = −3.34, T = 15). SUMMARY: 2/7 COUNTRIES COINTEGRATED; LOW POWER AT T = 15 (INDICATIVE, TO BE READ ALONGSIDE THE PANEL TESTS: KAO P = 0.013; WESTERLUND [39] P = 0.030)

Country	N	t ADF	5% CV	Conclusion
Belgium	15	−2.311	−3.34	No cointegration
Estonia	15	−2.761	−3.34	No cointegration
Greece	15	−0.487	−3.34	No cointegration
Hungary	15	−3.589	−3.34	Cointegration [5%]
Netherlands	15	−3.150	−3.34	Cointegration [10%]
Portugal	15	−1.442	−3.34	No cointegration
Slovenia	15	−2.304	−3.34	No cointegration

K. Half-Panel Jackknife Bias Correction for the Persistence Coefficient

TABLE XXI. HALF-PANEL JACKKNIFE (DHAENE AND JOCHMANS [33]) ON THE DYNAMIC FE PERSISTENCE COEFFICIENT (SUB-PANELS 2011-2017 AND 2018-2024; $\rho_{HPJ} = 2\rho_{FE} - (\rho_1+\rho_2)/2$). THE CORRECTION CONFIRMS THE EXPECTED DOWNWARD DIRECTION OF THE NICKELL BIAS BUT IS IMPRECISE AT N = 7 (A COUNTRY-CLUSTER BOOTSTRAP 95% INTERVAL SPANS [0.10, 1.09], REACHING THE NON-STATIONARY REGION); NO HALF-LIFE IS DERIVED FROM IT. THE INTERVAL COMES FROM THE COUNTRY-CLUSTER (PAIRS) BOOTSTRAP, TREATED AS UNRELIABLE AT N = 7 (SECTION IV-B) AND USED ONLY TO ILLUSTRATE IMPRECISION; THE EX-COVID FIGURE EXCLUDES 2020-2021. HERE $\rho_{FE} = 0.620$ IS THE LSDV ESTIMATE RECOMPUTED ON THE JACKKNIFE COMPLETE-CASE SAMPLE, MARGINALLY BELOW THE HEADLINE $\rho = 0.629$ OF TABLE 2 (WITHIN ESTIMATOR ON THE FULL DYNAMIC SAMPLE); ρ_{FE} UNDER WCB INFERENCE REMAINS THE MAIN RESULT

Estimate	ρ
ρ_{FE} (full sample)	0.620
sub-panel 2011-2017	0.552
sub-panel 2018-2024	0.363
ρ_{HPJ} (bias-corrected)	0.781
ρ_{HPJ} (ex-COVID)	0.938
95% bootstrap CI	[0.10, 1.09]

L. Predetermined (Lagged) Normalization of the Guarantee Ratio

In response to the suggestion of normalizing guarantees by a lagged stock rather than by GDP, and in the absence of an outstanding SME-loan-stock series in the dataset, the closest available predetermined scaling uses lagged new lending. The contemporaneous lagged-flow ratio G_t / $LOANNEW_{t-1}$ carries a positive coefficient that is only marginal under the Webb WCB ($\beta = +0.44$, $p_{WCB} = 0.082$) and becomes significant solely once a $1/LOANNEW_{t-1}$ control is added ($\beta = +0.47$, $p_{WCB} = 0.014$). This fragile significance reflects contemporaneous overlap between guaranteed and new loans: when the ratio is made fully predetermined, G_{t-1} / $LOANNEW_{t-2}$, the coefficient collapses and loses all significance ($\beta = -0.04$, $p_{WCB} = 0.88$). A clean test of the lagged-stock normalization would require the outstanding SME-loan-stock series, which was not available. Under the GDP normalization retained in the main text the ratio is not significant, whether contemporaneous ($\beta = +3.01$, $p_{WCB} = 0.40$) or lagged/predetermined ($guarantees_{t-1}/GDP_{t-1}$: $\beta = +3.12$, $p_{WCB} = 0.10$). The predetermined checks therefore confirm that the main result is not an artifact of contemporaneous overlap.

M. Size-Calibrated Power Analysis

TABLE XXII. POWER OF THE WEBB WCB TEST OF H0: BETA_RATIO = 0, ON THE ESTIMATED DESIGN (RECURSIVE DYNAMIC DATA GENERATION; B = 999; 1000 REPLICATIONS). THE TRUE EFFECT IS EXPRESSED IN LOG POINTS OF NEW LENDING PER ONE-STANDARD-DEVIATION RISE IN THE GUARANTEES/GDP RATIO. 'NOMINAL' USES THE 5% THRESHOLD; 'SIZE-CALIBRATED' USES THE SIMULATION-CALIBRATED CRITICAL VALUE ALPHA* = 0.087 (EMPIRICAL SIZE 0.030 AT THE NOMINAL 5% LEVEL). THE MINIMUM EFFECT DETECTABLE AT 80% POWER IS ABOUT 0.12 LOG POINTS PER STANDARD DEVIATION (~13%), WHEREAS THE ESTIMATED EFFECT IS ABOUT 0.047 (~5%). THE RATIO'S STANDARD DEVIATION AND THE OBSERVED EFFECT ARE THOSE OF THE POWER-ANALYSIS (LAG-COMPLETE) SAMPLE; ALPHA* IS SPECIFIC TO THIS TEST AND IS NOT APPLIED TO THE OTHER COEFFICIENTS

Effect (log pts / +1 SD)	Power nominal 5%	Power size-calibrated
0	0.03	0.05
0.05	0.15	0.30
0.10	0.51	0.69
0.15	0.75	0.94
0.20	0.84	0.98
0.30	0.92	1.00
0.50	1.00	1.00

N. Governance in First Differences (Robustness to the I(1) Classification)

TABLE XXIII. DYNAMIC FE (TWO-WAY, HC1-CLUSTERED, WEBB WCB) WITH GOVERNANCE ENTERED IN FIRST DIFFERENCES (98 OBS, 7 COUNTRIES). PERSISTENCE IS NOMINALLY WCB-SIGNIFICANT ($p_{WCB} = 0.026$) BUT SHOULD BE READ WITH CAUTION AT SEVEN CLUSTERS; Δ GOVERNANCE IS NOT SIGNIFICANT ($p_{WCB} = 0.33$). THE Δ-GOVERNANCE COEFFICIENT MEASURES THE EFFECT OF ANNUAL CHANGES IN GOVERNANCE — SMALL AND NOISY IN THE WGI SERIES — NOT OF ITS LEVEL. THE GUARANTEE RATIO REMAINS NON-SIGNIFICANT.

Variable	β	p_{WCB}
LOANNEW_LOG_LAG (ρ)	+0.673	0.026
LOANGOVGUARANTEES_RATIO	+1.786	0.585
INTERESTRATESPREAD	−0.011	0.930
NPLOUTSTANDING	−0.006	0.157
GDPGROWTH	−0.003	0.523
INFLATIONRATE	+0.002	0.899
Δ GOVERNANCEINDEX (first diff.)	+0.173	0.330