

Supplementary Materials

Does the Internet Replace Guanxi? Relational Investment and Income Quality in Rural China

Supplementary File S1. Additional Empirical Procedures

S1. Robustness and Alternative Index Procedures

The baseline specification is re-estimated under four variants that alter the sample or the inference procedure: after 1% winsorization of the continuous variables, after excluding municipalities, after excluding households with zero gift expenditure, and with standard errors clustered at the county level. In addition, a 500-draw placebo exercise benchmarks the estimated interaction against a random-reassignment distribution. The income-quality index is also reconstructed with alternative aggregation weights—entropy weighting, CRITIC weighting, principal-component weighting, and random weighting—applied to the second-stage aggregation across the four dimensions (adequacy, structure, stability, and cost). These alternative weighting methods modify only the second-stage aggregation; the within-dimension construction of the component indicators is unchanged. For the random-weighting exercise, 2,000 weight vectors are drawn from a Dirichlet(1,1,1,1) distribution with a fixed seed of 20260910. The corresponding estimates are reported in Table S2.

As an alternative measure of relational investment, we use the ratio of household gift expenditure to household consumption. We re-estimate the interaction using both OLS and a supplementary 2SLS specification; the corresponding results are reported in Panel C of Table S2. The 2SLS specification uses the same village-level leave-one-out peer instruments as the baseline, instrumenting the gift-expenditure ratio, Internet use, and their interaction jointly with the peer means of relational investment, Internet use, and their interaction, centered on the estimation sample.

S2. Marginal Effects and Heterogeneity Procedures

The marginal effect of relational investment is computed from the interaction model, in which income quality is regressed on relational investment, household Internet use, and their interaction. The marginal effect is evaluated at the observed values of household Internet use, and 95% confidence intervals are constructed from the estimated covariance matrix of the coefficients. The resulting curve is shown in Figure 2.

The baseline specification is estimated separately across predefined household or regional groups. Formal differences between groups are assessed through interaction-based tests and are reported in Table S4; subgroup coefficients on their own do not establish that the groups differ. The heterogeneity grouping variables are operationalized as follows.

Livelihood structure. Households are classified as pure farming (engaged in agriculture without off-farm work), part-time farming (engaged in agriculture with some off-farm work), or non-farming (not engaged in agriculture).

Clan-network density. The mean gift expenditure of each village is compared with the median of village means in the same survey year; households in villages at or above the annual median form the high-density group, and the remainder form the low-density group.

Tie strength. Using each household's average standardized log gift expenditure and average standardized log telecommunication expenditure, households for which gift expenditure dominates are classified as strong-tie households and the remainder as weak-tie households.

Digital information-use indicator. The indicator counts how many of the three economically oriented Internet activities recorded in the CFPS (learning, work, and business) the household uses at a non-zero, non-never frequency; households with two or more such activities form the high group and the remainder form the low group. The indicator is a coarse behavioural measure of the breadth of digital engagement rather than a measure of digital skill, and the corresponding subgroup results are treated as exploratory rather than as a test of H3.

The regional analysis follows the conventional division of Chinese provinces into eastern, central, and western China. Formal between-group tests for selected comparisons—clan-network density, region, livelihood structure, and digital information use—are reported in Table S4.

S3. Purpose-Specific Internet-Use Measures

The purpose-specific measures are constructed from the CFPS Internet-use module, which records five purposes of Internet use: learning, work, socializing, entertainment, and business. For each purpose, respondents report the frequency of Internet use on a 0–7 scale; the scale is reverse-coded so that higher values indicate more frequent use, and the reverse-coded value is averaged across adult household members. For households in which no adult reports using the Internet, the purpose-specific measure is set to zero. Each purpose-specific measure is then standardized to one standard deviation. The generic household Internet-use share is entered in its original proportion units (i.e., it is not standardized) and is mean-centered as in the baseline interaction specification; the final reference row of Table 6 reports the generic Internet-use measure and its interaction after scaling the measure to one standard deviation for comparison.

S4. Software, Inference, and Implementation Details

All estimations were conducted in Stata 15. The fixed-effects models were estimated with the user-written command `reghdfe`, with household fixed effects and county $\times$ year fixed effects absorbed by the within-transformation. The instrumental-variable exercises were implemented with `ivreghdfe` and `ivreg2`, with `ranktest` used for weak-instrument diagnostics. Standard errors are clustered at the village level in the baseline and subgroup specifications; the robustness analysis additionally reports standard errors clustered at the county level.

Table S1. Sensitivity Analyses for Potential Endogeneity

Identification strategy	RI $\times$ Internet Use	Identification information
Combined leave-one-out IV	−0.0876* (−1.71)	KP rk Wald F = 1.301; KP rk LM = 3.267 (p = 0.071); SW F = 7.54 / 11.17 / 5.32
Wooldridge nonlinear-IV specification	−0.0624*** (−3.04)	Nonlinear generated-instrument specification
Partial-endogeneity specification	−0.0425* (−1.80)	Internet use treated as conditionally exogenous
Relaxed-FE specification	−0.0585 (−0.93)	Household FE + province $\times$ year FE

Note: The dependent variable is the equal-weighted income-quality index. RI denotes relational investment. In the combined leave-one-out IV specification, RI, Internet use, and their interaction are jointly instrumented using the corresponding leave-one-out village-year peer measures. The reported SW F statistics refer, respectively, to RI, Internet use, and RI $\times$ Internet Use. The partial-endogeneity specification treats Internet use as conditionally exogenous, while the relaxed-FE specification replaces county-by-year with province-by-year fixed effects. Standard errors are clustered at the village level, and t-statistics are reported in parentheses. Significance is denoted by *** p < 0.01, ** p < 0.05, and * p < 0.10. These exercises are reported as supplementary sensitivity checks rather than as a causal identification strategy: the combined instruments are weakly identified, and the alternative specifications rest on assumptions that cannot be tested with these data.

Table S2. Robustness Checks

Specification / Measure	RI × Internet Use	Additional Information
A. Alternative samples and inference procedures		
Baseline	−0.0033*** (−2.67)	—
1% winsorization	−0.0032*** (−2.65)	—
Excluding municipalities	−0.0036*** (−2.88)	—
County-clustered standard errors	−0.0033** (−2.46)	—
Excluding zero-gift households	−0.0088*** (−2.84)	—
Placebo test, 500 draws	—	Only 1.0% of placebo draws have t exceeding the baseline value
B. Alternative income-quality construction		
Entropy weighting	−0.0011 (−1.41)	Weights: 0.023 / 0.143 / 0.820 / 0.015
CRITIC weighting	−0.0048*** (−2.70)	Weights: 0.242 / 0.433 / 0.075 / 0.250
First principal component	−0.0052*** (−2.80)	Loadings: 0.410 / 0.418 / −0.032 / 0.204
Random weights, 2,000 draws	Median = −0.0032	100% negative; 81.5% significant at 5%
C. Alternative relational-investment measure		
Gift-expenditure ratio × Internet use, OLS	0.0038 (0.24)	Not statistically significant
Gift-expenditure ratio × Internet use, 2SLS	−1.2005 (−1.00)	First-stage F = 3.6; weak identification

Note: The dependent variable in the baseline specification is the equal-weighted income-quality index. Dimension-specific indices are constructed using entropy weighting internally; the alternative weighting schemes modify only the second-stage aggregation across adequacy, structure, stability, and cost. All regressions include the baseline controls, household fixed effects, and county × year fixed effects unless otherwise stated. Standard errors are clustered at the village level except in the county-clustered specification. t-statistics are reported in parentheses. Significance is denoted

by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$. The entropy-weighted aggregate is attenuated because stability, for which the dimension-specific interaction is close to zero, receives 82.0% of the aggregate weight. For the random-weighting specification, 2,000 weight vectors were drawn from a Dirichlet(1,1,1,1) distribution with a fixed seed of 20260910.

Table S3. Regional Patterns of the Relational-Investment $\times$ Internet-Use Association

Region	N	Share of sample (%)	Mean IQ index	Mean Internet-use share	Mean ln(gift expenditure)	RI	Internet use	RI $\times$ Internet use
Eastern	4,747	35.8	0.5383	0.3012	6.5750	0.0023*** (3.59)	0.0312*** (4.59)	−0.0042*** (−2.88)
Central	3,945	29.7	0.5448	0.3088	6.9087	0.0032*** (3.78)	0.0323*** (4.58)	−0.0038* (−1.75)
Western	4,581	34.5	0.5335	0.2920	6.8246	0.0023** (2.16)	0.0253*** (3.15)	−0.0014 (−0.42)

Note: Regional groups follow the conventional division of Chinese provinces into eastern, central, and western China. For each region the coefficient on relational investment (RI), on Internet use, and on their interaction is estimated separately, with the full control set, household fixed effects, and county $\times$ year fixed effects; standard errors are clustered at the village level. IQ denotes the equal-weighted income-quality index. Because formal between-group tests are not uniformly statistically significant, the regional pattern should be read as suggestive rather than as evidence of established regional heterogeneity. t-statistics are reported in parentheses. Significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Table S4. Formal Between-Group Difference Tests

Grouping dimension	Equal-weighted index	Entropy-weighted index	N
Clan-network density (high vs. low)	0.0012 (0.45)	0.0006 (0.43)	13,273
Region (western vs. non-western)	0.0023 (0.64)	0.0008 (0.44)	13,273
Livelihood structure (non-farming vs. other)	−0.0005 (−0.22)	−0.0009 (−0.58)	13,178
Digital information use (high vs. low)	−0.0032 (−0.97)	−0.0043** (−2.01)	13,178

Note: Each cell reports the coefficient on the triple interaction between relational investment, household Internet use, and the indicated group dummy, with t-statistics in parentheses, estimated separately on the equal-weighted and entropy-weighted income-quality indices with the full specification (baseline controls, household fixed effects, and county $\times$ year fixed effects) and standard errors clustered at the village level. The group dummies compare high versus low clan-network density, western versus non-western provinces, non-farming versus farming and part-time livelihoods, and high versus low digital information use, so that each coefficient measures how much larger or smaller the relational-investment $\times$ Internet-use interaction is in the indicated group. Only the digital-information-use comparison under the entropy-weighted index is statistically significant; none of the other between-group differences is statistically significant, which is why the between-group comparisons in the main text are described as descriptive rather than as established heterogeneity. Significance is denoted by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.